(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(72) Inventors; and
(75) Inventors/Applicants (for US only): HOOYKAAS, Paul, Jan, Jacob [NL/NL]; Floris Versterlaan 12, NL-2343 RS Oegstgeest (NL). VAN ATTIKUM, Haico [NL/NL]; Boerhaavelaan 158, NL-2334 EV Leiden (NL). BUN-DOCK, Paul [GB/NL]; Mijehof 345, NL-1106 HJ Amsterdam (NL).

(54) Title: NUCLEIC ACID INTEGRATION IN EUKARYOTES

(57) Abstract: The invention relate to the field of molecular biology and cell biology. It particularly relates to methods to direct integration of a nucleic acid of interest towards homologous recombination and uses thereof. The present invention discloses factors involved in integration of a nucleic acid by illegitimate recombination which provides a method to direct integration of a nucleic acid of interest to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in a eukaryote with a preference for non-homologous recombination comprising steering an integration pathway towards homologous recombination. Furthermore, the invention provides a method to direct integration of a nucleic acid of interest to a sub-telomeric and/or telomeric region in a eukaryote with a preference for non-homologous recombination.

Title: Nucleic acid integration in eukaryotes

The invention relates to the field of molecular biology and cell biology. It particularly relates to methods to direct integration towards homologous

5   recombination and uses thereof. Several methods are known to transfer nucleic acids to, in particular, eukaryotic cells. In some methods the nucleic acid of interest is transferred to the cytoplasm of the cell, in some the nucleic acid of interest is integrated into the genome of the host. Many different vehicles for transfer of the nucleic acid are known. For different kinds of cells, different

10   systems can be used, although many systems are more widely applicable than just a certain kind of cells. In plants, e.g., a system based on *Agrobacterium tumefaciens* is often applied. This system is one of the systems that is used in a method according to the invention.

The soil bacterium *Agrobacterium tumefaciens* is able to transfer part of

15   its tumor-inducing (Ti) plasmid, the transferred (T-) DNA, to plant cells. This results in crown gall tumor formation on plants due to expression of *onc*-genes, which are present on the T-DNA. Virulence (*vir*) genes, located elsewhere on the Ti-plasmid, mediate T-DNA transfer to the plant cell. Some Vir proteins accompany the T-DNA during its transfer to the plant cell to protect the T-

20   DNA and to mediate its transfer to the plant nucleus. Once in the plant nucleus, the T-DNA is integrated at a random position into the plant genome (reviewed by (Hooykaas and Beijersbergen, 1994), (Hansen and Chilton 1999). Removal of the *onc*-genes from the T-DNA does not inactivate T-DNA transfer. T-DNA, disarmed in this way, is now the preferred vector for the genetic

25   modification of plants.

Although much is known about the transformation process, not much is known about the process by which the T-DNA is integrated into the plant genome. It is likely that plant enzymes mediate this step of the transformation process (Bundock at al. 1995). The integration pattern of T-DNA in

30   transformed plants has been extensively studied (Matsumoto at al. 1990) (Gheysen et al 1991) (Meyerhofer et al 1991). The results indicated that T-

DNA integrates via illegitimate recombination (IR) (also called non-homologous recombination, both terms may be used interchangeable herein), a process which can join two DNA molecules that share little or no homology (here the T-DNA and plant target DNA). Even T-DNA molecules in which a large segment of homologous plant DNA was present, integrated mainly by IR and only with very low frequency ($1:10^4$–10^5) by homologous recombination (HR) (Offringa et al. 1990).

Recently, it was shown that *Agrobacterium* is not only able to transfer its T-DNA to plant cells, but also to other eukaryotes, including the yeast *S.cerevisiae* (Bundock et al. 1995) and a wide variety of filamentous fungi (de Groot et al. 1998). In *S.cerevisiae*, T-DNA carrying homology with the yeast genome integrates via HR (Bundock et al. 1995). However, T-DNA lacking any homology with the *S.cerevisiae* genome becomes integrated at random positions in the genome by the same IR process as is used in plants (Bundock and Hooykaas 1996). Apparently, eukaryotic cells have at least two separate pathways (one via homologous and one via non-homologous recombination) through which nucleic acids (in particular of course DNA), can be integrated into the host genome. The site of integration into a host cell genome is important with respect to the likelihood of transcription and/or expression of the integrated nucleic acid. The present invention provides methods and means to direct nucleic acid integration to a pre-determined site through steering integration towards the homologous recombination pathway. The present invention arrives at such steering either by enhancing the HR pathway, or by inhibiting (meaning reducing) the IR pathway.

Host factors involved in the integration of nucleic acid by IR have so far not been identified. The present invention discloses such factors which enables the design of methods for their (temporary) inhibition, so that integration of nucleic acid by IR is prevented or more preferably completely inhibited, shifting the integration process towards HR and facilitating the isolation of a host cell with nucleic acid integrated by HR at a predetermined site. This is

extremely important, since there is no method available yet for easy and
precise genetic modification of a host cell using HR (gene targeting). Of course
the actual site of integration is then determined by homology of the nucleic
acid of interest with said site.

5   In a first embodiment the invention provides a method to direct nucleic
acid integration of a nucleic acid of interest to a pre-determined site, whereby
said nucleic acid has homology at or around the said pre-determined site, in a
eukaryote with a preference for non-homologous recombination comprising
steering an integration pathway towards homologous recombination.

10   Preferably, such a method comprises at least the steps of introducing said
nucleic acid of interest to a cell of said eukaryote, for example by the process of
transformation or electroporation, and integration of said nucleic acid in the
genetic material of said cell. Integration is a complex process wherein a nucleic
acid sequence becomes part of the genetic material of a host cell. One step in

15   the process of nucleic acid integration is recombination; via recombination
nucleic acid sequences are exchanged or inserted and the introduced nucleic
acid becomes part of the genetic material of a host cell. In principle two
different ways of recombination are possible: homologous and illegitimate or
non-homologous recombination. Most (higher) eukaryotes do not or at least not

20   significantly practise homologous recombination although the essential
proteins to accomplish such a process are available. One reason for this
phenomenon is that frequent use of homologous recombination in (higher)
eukaryotes could lead to undesirable chromosomal rearrangements due to the
presence of repetitive nucleic acid sequences. To accomplish homologous

25   recombination via a method according to the invention, it is important to
provide a nucleic acid which has homology with a pre-determined site. It is
clear to a person skilled in the art that the percentage of homology and the
length of (a) homologous region(s) play(s) an important role in the process of
homologous recombination. The percentage of homology is preferably close to

30   100%. A person skilled in the art is aware of the fact that lower percentage of

4

homology are also used in the field of homologous recombination, but dependent on, for example, the regions of homology and their overall distribution, can lead to a lower efficiency of homologous recombination but are still useful and therefore included in the present invention. Furthermore,

5   the length of a (nearly) homologous region is approximately 3 kb which is sufficient to direct homologous recombination. At least one homologous region is necessary for recombination but more preferably 2 homologous regions flanking the nucleic acid of interest are used for targeted integration. The researcher skilled in the art knows how to select the proper percentage of

10   homology, the length of homology and the amount of homologous regions. By providing such a homology a nucleic acid is integrated at every desired position within the genetic material of a host cell. It is clear to a person skilled in the art that the invention as disclosed herein is used to direct any nucleic acid (preferably DNA) to any pre-determined site as long as the length of

15   homology and percentage of homology are high enough to provide/enable homologous recombination. A pre-determined site is herein defined as a site within the genetic material contained by a host cell to which a nucleic acid with homology to this same site is integrated with a method according to the invention. It was not until the present invention that a nucleic acid is

20   integrated at every desired position and therefore a method according to the invention is applied, for example, to affect the gene function in various ways, not only for complete inactivation but also to mediate changes in the expression level or in the regulation of expression, changes in protein activity or the subcellular targeting of an encoded protein. Complete inactivation,

25   which can usually not be accomplished by existing methods such as antisense technology or RNAi technology (Zrenner et al, 1993), is useful for instance for the inactivation of genes controlling undesired side branches of metabolic pathways, for instance to increase the quality of bulk products such as starch, or to increase the production of specific secondary metabolites or to inhibit

30   formation of unwanted metabolites. A method according to the invention is

also used to inactivate genes controlling senescence in fruits and flowers or that determine flower pigments. Replacement of existing regulatory sequences by alternative regulatory sequences is used to alter expression of in situ modified genes to meet requirements, (e.g. expression in response to particular

5   physical conditions such as light, drought or pathogen infection, or in response to chemical inducers), or depending on the developmental state (e.g. in a storage organ, or in fruits or seeds) or on tissue or cell types. Also a method according to the invention is used to effectuate predictable expression of transgenes encoding novel products, for example by replacing existing coding

10 sequences of genes giving a desired expression profile by those for a desired novel product. For example to produce proteins of medicinal or industrial value in the seeds of plants the coding sequence of a strongly expressed storage protein may be replaced by that of the desired protein. As another example existing coding sequences are modified so that the protein encoded has

15 optimized characteristics for instance to make a plant herbicide tolerant, to produce storage proteins with enhanced nutritional value, or to target a protein of interest to an organelle or to secrete it to the extracellular space. As yet another example eukaryotic cells (including yeast, fungus, plant, mammalian cells or (non-human) animal cells) are provided with a gene

20 encoding a protein of interest integrated into the genome at a site which ensures high expression levels. As another example the nucleic acid of interest can be part of a gene delivery vehicle to deliver a gene of interest to a eukaryotic cell in vitro or in vivo. In this way a defect p53 can be replaced by an intact p53. In this way a tumoricidal gene is delivered to a pre-determined

25 site present only in e.g. proliferating cells, or present only in tumor cells, e.g. to the site where a tumor antigen is expressed from. Gene delivery vehicles are well known in the art and include adenoviral vehicles, retroviral vehicles, non-viral vehicles such as liposomes, etc. As another example the invention is used to produce transgenic organisms. Knock-out transgenics are already produced

30 by homologous recombination methods. The present invention improves the

6

efficiency of such methods. Also transgenics with desired properties are made. It is clear to a person skilled in the art that transgenics can for example be made by the use of *Agrobacterium* as gene delivery vehicle for plant (Vergunst et al., 1998), yeast (Bundock et al., 1995), fungus (de Groot et al., 1998) or

5   animal (Kunik et al., 2001) or by direct DNA delivery methods exemplified by but not restricted to electroporation for yeast (Gietz & Woods, 2001), plant (D'Halluin et al., 1992; Lin et al., 1997), fungus (Ozeki et al, 1994) and animal (Templeton et al., 1997), LiCl treatment for yeast (Schiestl et al., 1993), micro-injection for plant (Schnorf et al., 1991) and animal (Capecchi, 1980) and

10   "DNA whiskers" for plant (Kaeppler et al., 1990; Dunwell, 1999) or particle bombardment for plants and animals (Klein et al., 1992). It is furthermore clear that transgenic plants can be obtained via selective regeneration of transformed plant cells into a complete fertile plant (Vergunst et al., 1998) or via non-regenerative approaches by transforming germ line cells, exemplified

15   by but not restricted to dipping *Arabidopsis* flowers into an *Agrobacterium* suspension (Bechtold et al., 1993). It is also clear that transgenic animals can be obtained by transforming embryonic stem cells with one of the DNA delivery methods mentioned above (Hooper, 1992).

In another embodiment the invention provides a method to direct

20   nucleic acid integration to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in a eukaryote with a preference for non-homologous recombination comprising steering an integration pathway towards homologous recombination by providing a mutant of a component involved in non-homologous recombination. Methods to

25   identify components involved in non-homologous recombination are outlined in the present description wherein *S.cerevisiae* was used as a model system. To this end several yeast derivatives defective for genes known to be involved in various recombination processes were constructed and the effect of the mutations on T-DNA integration by either HR or IR was tested. The results as

30   disclosed herein show that the proteins encoded by *YKU70, RAD50, MRE11,*

XRS2, LIG4 and *SIR4* play an essential role in DNA integration by IR but not by HR. WO 00/12716 describes a maize Ku70 orthologue and suggests that "Control of homologous recombination or non-homologous end joining by modulating Ku provides the means to modulate the efficiency which

5   heterologous nucleic acids are incorporated into the genomes of a target plant cell." WO 00/68404 describes a maize Rad50 orthologue and suggest an analogous control for Rad50. Both patent applications do however not disclose, in contrast to the present patent application, that by preventing or more preferably completely inhibiting non-homologous recombination, for example

10   by providing a mutant of a component involved in non-homologous recombination or by inhibiting such a component, that the integration pathway is steered towards homologous recombination. It is clear to a person skilled in the art that different mutants of a component involved in non-homologous recombination exist. Examples are deletion mutants, knock-out

15   (for example via insertion) mutants or point mutants. Irrespective of the kind of mutant it is important that a component involved in non-homologous recombination is no longer capable or at least significantly less capable to perform it's function in the process of non-homologous recombination. As disclosed herein disruption of *YKU70, RAD50, MRE11, XRS2, LIG4* and *SIR4*

20   did not affect the frequency of DNA integration by HR, showing that these genes are not involved in DNA integration by HR, but only in DNA integration by IR. More over, in the wild-type yeast strain 85% of the integration events occurred by HR (37% by replacement and 63% by insertion) and 15% by IR. In contrast, integration occurred only by HR in yeast strains lacking *ku70* or *lig4*.

25   In *rad50* and *xrs2* mutant strains the T-DNA preferentially integrated by HR (92%) and 93% of these T-DNAs integrated by replacement and only 7% by insertion. Thus, the absence of a functional *rad50* or *xrs2* gene leads to a significantly increased frequency of replacement reactions.

In another embodiment the invention provides a method to direct

30   integration of a nucleic acid of interest to a sub-telomeric and/or telomeric

8

region in a eukaryote with a preference for non-homologous recombination by
providing a mutant of a component involved in non-homologous recombination.
A telomeric region is typically defined as a region containing repetitive
sequences which is located at the end of a chromosome. A sub-telomeric region
is typically defined as a region flanking the telomeric region. As an example it
is disclosed herein that in yeast strains carrying disruptions of *RAD50,
MRE11* or *XRS2* the distribution of integrated DNA copies is altered when
compared to wildtype. DNA becomes preferentially integrated in telomeres or
subtelomeric regions in the *rad50, mre11* and *xrs2* mutants. A great advantage
of integration of DNA copies in telomeres or subtelomeric regions instead of
integration elsewhere in the genomic material is that there is no danger for
host genes being mutated or inactivated by a DNA insertion. When in plants
deficient for *RAD50, MRE11* or *XRS2* DNA copies also integrate into telomeres
or subtelomeric regions, such plants are used for (sub)telomeric targeting of T-
DNA in transformation experiments to prevent additional insertion mutations
from random T-DNA integration into the plant genome.

In yet another embodiment the invention provides a method to
direct nucleic acid integration to a pre-determined site, whereby said nucleic
acid has homology at or around the said pre-determined site, in a eukaryote
with a preference for non-homologous recombination comprising steering an
integration pathway towards homologous recombination by partially or more
preferably completely inhibiting a component involved in non-homologous
recombination. Partial or complete inhibition of a component involved in non-
homologous recombination is obtained by different methods, for example by an
antibody directed against such a component or a chemical inhibitor or a
protein inhibitor or peptide inhibitor or an antisense molecule or an RNAi
molecule. Irrespective of the kind of (partial or more preferably complete)
inhibition it is important that a component involved in non-homologous
recombination is no longer capable or at least significantly less capable to
perform it's function in the process of non-homologous recombination. In yet

another embodiment the invention provides a method to direct integration of a nucleic acid of interest to a sub-telomeric and/or telomeric region in a eukaryote with a preference for non-homologous recombination by partially or more preferably completely inhibiting a component involved in non-

5   homologous recombination. Preferably, said component involved in non-homologous recombination is *rad50*, *mre11* or *xrs2*.

In a preferred embodiment the invention provides a method to direct nucleic acid integration to a pre-determined site or to a sub-telomeric and/or telomeric region by providing a mutant of a component involved in non-

10   homologous recombination or by partially or more preferably completely inhibiting a component involved in non-homologous recombination wherein said component comprises *ku70*, *rad50*, *mre11*, *xrs2*, *lig4*, *sir4* or others such as *ku80* (Tacciole et al., 1994; Milne et al., 1996), *lif1* (Teo and Jackson, 2000; XRCC4 in human, see figure 6; Junop et al., 2000) and *nej1*, (Kegel et al.,

15   2001; Valencia et al., 2001). Components involved in non-homologous recombination are identified as outlined in the present description. The nomenclature for genes as used above is specific for yeast. Because the nomenclature of genes differs between organisms a functional equivalent or a functional homologue (for example *NBS1*, a human *xrs2* equivalent (Paull and

20   Gellert, 1999) and see for example figure 2 to 5) and/or a functional fragment thereof, all defined herein as being capable of performing (in function, not in amount) at least one function of the yeast genes *ku70*, *rad50*, *mre11*, *xrs2*, *lig4*, *sir4*, *ku80*, *lif1* or *nej1* are also included in the present invention. A mutant of a component directly associating with a component involved in non-

25   homologous recombination or (partial or complete) inhibition of a component directly associating with a component involved in non-homologous recombination is also part of this invention. Such a component directly associating with a component involved in non-homologous recombination is, for example, identified in a yeast two hybrid screening. An example of a

30   component directly associating with a component involved in non-homologous

recombination is *KU80*, which forms a complex with *KU70*. In a more preferred embodiment the invention provides a method to direct nucleic acid integration in yeast, fungus, plant or (non-human) animal (cells).

In another embodiment the invention provides a method to direct
5 nucleic acid integration to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in a eukaryote with a preference for non-homologous recombination comprising steering an integration pathway towards homologous recombination by transiently (partially or more preferably completely) inhibiting integration via non-
10 homologous recombination. In yet another embodiment the invention provides a method to direct integration of a nucleic acid of interest to a subtelomeric and/or telomeric region in a eukaryote with a preference for non-homologous recombination by transiently (partially or more preferably completely) inhibiting integration via non-homologous recombination. In a more preferred
15 embodiment, such a method is used for yeast, plant, fungus or (non-human) animal and the transient (partial or more preferably complete) inhibition is provided by a (preferably stably) inserted and expressed chimeric transgene that encodes a peptide inhibitory to one, some or all non-homologous recombination (NHR) enzymes fused to a nuclear localisation signal (Hanover,
20 1992; Raikhel, 1992) and the steroid-binding domain of a steroid receptor (Picard et al., 1988). The chimeric transgene is constructed in such a way, using either heterologous or non-heterologous promoter sequences and other expression signals, that it provides (stable) expression in the target cells or tissue for transformation. In the absence of the steroid hormone, the steroid-
25 binding domain binds to chaperone proteins, and thereby the fusion protein is retained in the cytoplasm. Upon treatment with the steroid hormone, the chaperones are released from the steroid-binding domain and the inhibitory peptide will enter the nucleus where it will interact with and inhibit the action of NHR enzymes. An example of an inhibitory peptide is a KU80 fragment that

imparts radiosensitivity to Chinese hamster ovary cells (Marangoni et al., 2000).

In a more preferred embodiment such a method is used for yeast, plant, fungus or a (non-human) animal and the transient (partial or more preferable complete) inhibition is provided by an *Agrobacterium* Vir-fusion protein capable of (partially or more preferably completely) inhibiting a component involved in non-homologous recombination or capable of (partially or more preferably completely) inhibiting a functional equivalent or homologue thereof or capable of (partially or more preferably completely) inhibiting a component directly associating with a component involved in non-homologous recombination. In an even more preferred embodiment such an *Agrobacterium* Vir fusion protein comprises VirF or VirE2. It was shown that the *Agrobacterium* VirF and VirE2 proteins are directly transferred from *Agrobacterium* to plant cells during plant transformation (Vergunst et al. 2000). To, for example, accomplish T-DNA integration by HR in plants, VirF fusion proteins containing for example a peptide inhibitor of IR in plant cells are introduced concomitantly with the targeting T-DNA. It has been reported that the C-terminal part (approximately 40 amino acids) of VirF or VirE2 is sufficient to accomplish transfer of T-DNA. A functional fragment and/or a functional equivalent of VirF or VirE is therefore also included in the present invention. Preferably, said nucleic acid of interest is delivered to a cell of said eukaryote by *Agrobacterium*.

In an even more preferred embodiment a component involved in non-homologous recombination comprises *ku70, rad50, mre11, xrs2, lig4, sir4, ku80, lif1* or *nej1* or functional equivalents or homologous thereof or associating components. The nomenclature for genes as used above is specific for yeast. Because the nomenclature of genes differs between organisms a functional equivalent or a functional homologue (see for example figure 2 to 5) and/or a functional fragment thereof, all defined herein as being capable of performing (in function, not in amount) at least one function of the yeast genes

12

ku70, *rad50*, *mre11*, *xrs2*, *lig4*, *sir4*, *ku80*, *lif1* or *nej1* are also included in the present invention. By transiently (partially or more preferably completely) inhibiting a component involved in non-homologous recombination a nucleic acid is integrated at any desired position without permanently modifying a

5   component involved in non-homologous recombination and preventing unwanted side effects caused by the permanent presence of such a modified component involved in non-homologous recombination.

Methods according to the present invention, as extensively but not limiting discussed above, are used in a wide variety of applications. One

10   embodiment of the present invention is the replacement of an active gene by an inactive gene according to a method of the invention. Complete inactivation, which can usually not be accomplished by existing methods such as antisense technology or RNAi technology, is useful for instance for the inactivation of genes controlling undesired side branches of metabolic

15   pathways, for instance to increase the quality of bulk products such as st:: or to increase the production of specific secondary metabolites or to inhibit formation of unwanted metabolites. Also to inactivate genes controlling senescence in fruits and flowers or that determine flower pigments. Another embodiment of the present invention is the replacement of an inactive gene by

20   an active gene. One example is the replacement of a defect p53 by an intact p53. Many tumors acquire a mutation in p53 during their development which renders it inactive and often correlates with a poor response to cancer therapy. By replacing the defect p53 by an intact p53, for example via gene therapy, conventional anti cancer therapy have better changes of succeeding. In yet

25   another embodiment of the invention a therapeutic proteinaceous substance is integrated via a method of the invention. In this way a tumoricidal gene ise delivered to a pre-determined site present only in e.g. proliferating cells, or present only in tumor cells, e.g. to the site where a tumor antigen is expressed from. In yet another embodiment the invention provides a method to introduce

30   a substance conferring resistance for an antibiotic substance to a cell according

to a method of the invention. Also a method according to the invention is used
to confer a desired property to an eukaryotic cell. In a preferred embodiment a
gene delivery vehicle is used to deliver a desired nucleic acid to a pre-
determined site. Gene delivery vehicles are well known in the art and include
5   adenoviral vehicles, retroviral vehicles, non-viral vehicles such as liposomes,
etc.. In this way, a for example, tumoricidal gene can be delivered to a pre-
determined site present only in e.g. proliferating cells, or present only in tumor
cells, e.g. to the site where a tumor antigen is expressed from.

Furthermore a method according to the invention is used to improve
10   gene targeting efficiency. Such a method is used to improve for example the
gene targeting efficiency in plants. In plants transgenes integrate randomly
into the genome by IR (Mayerhof et al. 1991), (Gheysen et al. 1991). The
efficiency of integration by HR is very low, even when large stretches of
homology between the transgene and the genomic target site are present
15   (Offringa et al. 1990). Therefore, the efficiency of gene targeting using HR is
very low in plants. The results that are disclosed herein show how to improve
the gene targeting efficiency in plants. From the fact that T-DNA integration
by IR is strongly reduced in *KU70, RAD50, MRE11, XRS2, LIG4* and *SIR4*
deficient yeast strains and T-DNA integration by HR is not affected in such
20   strains, T-DNA integration by HR is more easily obtained in plants, deficient
for either of these genes. Recently, we have cloned a *KU70* homologue of
Arabidopsis thaliana (see figure 2, Bundock 2000, unpublished data). *RAD50,*
MRE11 and *LIG4* homologues have already been found in *A.thaliana*
(GenBank accession numbers AF168748, AJ243822 and AF233527,
25   respectively, see also figure 3, 4 and 5 (Hartung and Puchta 1999). Currently,
screenings are being performed to find plants carrying a T-DNA inserted in
AtMRE11, AtKU70 or *AtLIG4*. These knockout plants are used to test whether
T-DNA integration by IR is reduced and integration by HR is essentially
unaffected, thereby facilitating the detection of T-DNA integration by HR.

14

Furthermore, the invention provides a method to direct integration of a nucleic acid of interest to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in a eukaryote with a preference for non-homologous recombination, comprising steering an

5   integration pathway towards homologous recombination, wherein said nucleic acid sequence of interest is essentially replacing a sequence within said eukaryote. As disclosed herein within the experimental part, in the wild-type yeast strain 85% of the integration events occurred by HR (37% by replacement and 63% by insertion) and 15% by IR. In contrast, integration

10   occurred only by HR in yeast strains lacking *ku70* or *lig4*. In *rad50* and *xrs2* mutant strains the T-DNA preferentially integrated by HR (92%) and 93% of these T-DNAs integrated by replacement and only 7% by insertion. Thus, the absence of a functional *rad50* or *xrs2* gene leads to a significantly increased frequency of the desired replacement reactions.

15

The invention will be explained in more detail in the following description, which is not limiting the invention.

EXPERIMENTAL PART

Yeast strains.

The yeast strains that were used are listed in Table 1. Yeast mutants

5   isogenic to the haploid YPH250 strain were constructed using the one-step

disruption method (Rothstein 1991). A 1987 bp fragment from the *YKU70*

locus was amplified by PCR using the primers hdf1p1 5'-

GGGATTGCTTTAAGGTAG-3' and hdf1p2 5'-CAAATACCCTACCCTACC-3'.

The PCR product was cloned into pT7Blue (Novagen) to obtain

10   pT7Blue*YKU70*. A 1177 bp *EcoRV/Hin*dIII fragment from the *YKU70* ORF

was replaced by a 2033 bp *Hin*dIII/*Sma*I *LEU2* containing fragment from

pJJ283 (Jones and Prakash 1990), to form pT7Blue*YKU70::LEU2*. In order to

obtain YKU70 disruptants Leu[+] colonies were selected after transformation of

YPH250 with a 2884 bp *Nde*I/*Sma*I fragment from pT7Blue*YKU70::LEU2*.

15   The Expand™ High Fidelity System (Boehringer Mannheim) was used

according to the supplied protocol to amplify a 3285 bp fragment from the

LIG4 locus with primers

dnl4p1 5'-CGTAAGATTCGCCGAGTATAG-3' and

dnl4p2 5'-CGTTTCAAATGGGACCACAGC-3'. The PCR product was cloned

20   into pGEMT (Promega), resulting in pGEMT*LIG4*. A 1326 bp *Bam*HI/*Xho*I

fragment from pJJ215 (Jones and Prakash 1990) containing the *HIS3* gene

was inserted into the *Bam*HI and *Xho*I sites of pIC20R, resulting in

pIC20R*HIS3*. A 782 bp *Eco*RI fragment from the *LIG4* ORF was replaced with

a 1367 bp EcoRI *HIS3* containing fragment from pIC20R*HIS3* to construct

25   pGEMT*LIG4::HIS3*. In order to obtain LIG4 disruptants His[+] colonies were

selected after transformation of YPH250 with a 3854 bp *Nco*I/*Not*I fragment

from pGEMT*LIG4::HIS3*. In order to obtain RAD50 disruptants YPH250 was

transformed with a *Eco*RI/*Bgl*II fragment from pNKY83 and Ura[+] colonies

were selected (Alani et al 1989). A *rad50::hisG* strain was obtained by

30   selecting Ura[-] colonies on selective medium containing 5-FOA. Similarly

RAD51 disruptants were obtained were obtained after transformation of
YPH250 with a *RAD51::LEU2 XbaI/PstI* fragment from pDG152 and selection
of Leu[+] colonies (Schiestl et al 1994). The *TRP1* marker in pSM21 (Schild et al.
1983) was replaced with a *BglII/XbaI LEU2* containing fragment from pJJ283

5   (Jones and Prakash, 1990). This resulted in pSM21*LEU2*. Leu[+] RAD52
disruptant colonies were selected after transformation of YPH250 with the
RAD52::LEU2 BamHI fragment from pSM21*LEU2*. Disruption constructs
were transformed to YPH250 by the lithium acetate transformation method as
described (Gietz et al. 1992; Schiestl et al. 1993). Disruption of *YKU70, LIG4,*

10 *RAD50, RAD51* and *RAD52* was confirmed by PCR and Southern blot
analysis.

Table 1: Yeast strains

Strain	Genotype	Reference
YPH250	*MATα, ura3-52, lys2-801, ade2-101, trp1-Δ1, his3-Δ200, leu2-Δ1*	(Sikorski and Hieter 1989)
YPH250*rad51*	*MATα, ura3-52, lys2-801, ade2-101, trp1-Δ1, his3-Δ200, leu2-Δ1, rad51::LEU2*	This study
YPH250*rad52*	*MATα, ura3-52, lys2-801, ade2-101, trp1-Δ1, his3-Δ200, leu2-Δ1, rad52::LEU2*	This study
YPH250*yku70*	*MATα, ura3-52, lys2-801, ade2-101, trp1-Δ1, his3-Δ200, leu2-Δ1, yku70::LEU2*	This study
YPH250*rad50*	*MATα, ura3-52, lys2-801, ade2-101, trp1-Δ1, his3-Δ200, leu2-Δ1,rad50::hisG*	This study
YPH250*lig4*	*MATα, ura3-52, lys2-801, ade2-101, trp1-Δ1, his3-Δ200, leu2-Δ1, lig4::HIS3*	This study
JKM115	*Δho, Δhml::ADE1, MATα, Δhmr::ADE1, ade1, leu2-3,112, lys5, trp1::hisG, ura3-52*	(Moore and Haber 1996)
JKM129	*Δho, Δhml::ADE1, MATα, Δhmr::ADE1, ade1, leu2-3,112, lys5, trp1::hisG, ura3-52, xrs2::LEU2*	(Moore and Haber 1996)
JKM138	*Δho, Δhml::ADE1, MATα, Δhmr::ADE1, ade1, leu2-3,112, lys5, trp1::hisG, ura3-52, mre11::hisG*	(Moore and Haber 1996)
YSL204	*Δho, HMLα, MATα, HMRα, ade1-100, leu2-3,112, lys5, trp1::hisG, ura3-52, hisG'-URA3-hisG', sir4::HIS3*	(Lee et al. 1999)

18

Construction of binary vectors.

To construct pSDM8000 a 1513 bp *Pvu*II/*Eco*RV fragment carrying the *KanMX* marker was obtained from pFA6a (Wach et al. 1994) and was ligated into the unique *Hpa*I site of pSDM14 (Offringa 1992). pSDM8001 was made in

5   three cloning steps. A 1476 bp *Bam*HI/*Eco*RI fragment carrying the *KanMX* marker was obtained from pFA6a and ligated into *Bam*HI and *Eco*RI digested pIC20H to form pIC20H*kan*MX. The KanMX marker was inserted between the *PDA1* flanks by replacement of a 2610 bp *Bgl*II fragment from pUC4E1α10 (Steensma et al. 1990) with a 1465 *Bgl*II fragment from pIC20H*kan*MX. A

10   3721 bp *Xho*I/*Kpn*I fragment from this construct was inserted into the *Xho*I and *Kpn*I sites of pSDM14. The binary vectors pSDM8000 and pSDM8001 were introduced into *Agrobacterium tumefaciens* LBA1119 by electroporation (den Dulk-Ras and Hooykaas 1995).

15   **Cocultivations / T-DNA transfer experiments.**

Cocultivations were performed as described earlier with slight modifications (Bundock et al. 1995). *Agrobacterium* was grown overnight in LC medium. The mix of *Agrobacterium* and *S. cerevisiae* cells was incubated for 9 days at 20°C. G418 resistant *S.cerevisiae* strains were selected at 30°C on

20   YPAD medium containing geneticin (200 µg/ml) (Life Technologies/Gibco BRL).

Vectorette PCR.

Chromosomal DNA was isolated using Qiagen's Genomic Tips G/20 per

25   manufacters protocol. 1-2 µg of Genomic DNA was digested with *Eco*RI, *Cla*I, *Pst*I or *Hin*dIII and run on a 1% TBE-gel. Non-radioactive Southern blotting was performed. The membrane was hybridized with a digoxigenine-labeled kanMX probe to determine the size of T-DNA/genomic DNA fragments (*Eco*RI and *Cla*I for RB containing fragments and *Pst*I and *Hin*dIII for LB containing

30   fragments). The kanMX probe, a 792 bp internal fragment of the *KanMX*

marker, was made by PCR using primers kanmxp1 5'-
AGACTCACGTTTCGAGGCC-3' and kanmxp2 5'-
TCACCGAGGCAGTTCCATAG-3' and a Non-Radioactive DNA Labeling and
Detection kit (Boehringer Mannheim). The enzyme showing the smallest band

5   on blot was used for Vectorette PCR, in order to amplify the smallest junction
sequence of T-DNA and genomic DNA. Vectorette PCR was performed as
described
(http://genomewww.stanford.edu/group/botlab/protocols/vectorette.html). The
Expand™ High Fidelity System (Boehringer Mannheim) was used to amplify

10   fragments larger than 2.5 kb, whereas sTaq DNA polymerase (SphaeroQ) was
used for amplification of fragments smaller than 2.5 kb. Primer kanmxp3 5'-
TCGCAGGTCTGCAGCGAGGAGC-3' and kanmxp4 5'-
TCGCCTCGACATCATCTGCCCAG-3' were used to amplify RB/genomic DNA
and LB/genomic DNA junction sequences, respectively.

15

T7 DNA Polymerase sequencing.

Vectorette PCR products were cloned in pGEMTEasy (Promega) and
sequenced using the T7 polymerase sequencing kit (Pharmacia) according to
manufacturers protocol. In order to obtain sequences flanking the RB and LB,

20   primers kanmxp5 5'-TCACATCATGCCCCTGAGCTGC-3' and kanmxp4 were
used, respectively.

20

RESULTS

1. Binary vectors for T-DNA transfer to yeast.

It was previously demonstrated that *Agrobacterium tumefaciens* is able
5   to transfer its T-DNA not only to plants but also to another eukaryote, namely
the yeast *Saccharomyces cerevisiae* (Bundock et al. 1995). T-DNA carrying
homology with the yeast genome was shown to become integrated by
homologous recombination. T-DNA lacking any homology with the yeast
genome was integrated randomly into the genome by IR like in plants
10   (Bundock et al. 1995, (Bundock and Hooykaas 1996). The T-DNA used in these
experiments carried the *S.cerevisiae URA3* gene for selection of Ura+ colonies
after T-DNA transfer to the haploid yeast strain RSY12(*URA3Δ*). However, in
this system only yeast strains could be used in which the *URA3* gene had been
deleted to avoid homology between the incoming T-DNA and the *S.cerevisiae*
15   genome.

We wanted to setup a system in which T-DNA transfer to any yeast
strain could be studied. Therefore, two new binary vectors were constructed
using the dominant marker *kanMX* (Wach et al 1994)which confers resistance
against geneticin (G418). The T-DNA of pSDM8000 carries only the *KanMX*
20   marker. Since this *KanMX* marker consists of heterologous DNA, lacking any
homology with the *S.cerevisiae* genome, we would expect this T-DNA to
integrate by IR at a random position in the yeast genome. To be able to
compare this with T-DNA integration by homologous recombination
pSDM8001 was constructed. The T-DNA of pSDM8001 carries the *KanMX*
25   marker flanked by sequences from the *S.cerevisiae PDA1* locus. The *PDA1*
sequences have been shown to mediate the integration of T-DNA by HR at the
PDA1 locus on chromosome V (Bundock et al. 1995).

Cocultivations between *Agrobacterium* strains carrying pSDM8000 and
pSDM8001, respectively, and the haploid yeast strains YPH250 and JKM115,
30   respectively, were carried out as described in the experimental part. G418

resistant colonies were obtained at low frequencies for YPH250 (1.6×10^{-7}) and JKM115 (1.2×10^{-5}) after T-DNA transfer from pSDM8000 (Table 2). T-DNA transfer from pSDM8001 generated G418 resistant colonies at higher frequencies (2.4×10^{-5} for YPH250 and 1.8×10^{-4} JKM115, Table 2). The ratio

5   of homologous recombination versus illegitimate recombination is determined by comparing the frequencies of G418 resistant colonies obtained from cocultivations using either pSDM8001 or pSDM8000. This showed that a T-DNA from pSDM8001 was 150-fold more likely to integrate than a T-DNA from pSDM8000 in YPH250 (Table 2). A similar difference was previously seen

10   using T-DNAs with the *URA3* marker (Bundock and Hooykaas 1996). In contrast, T-DNA from pSDM8001 was only 16-fold more likely to integrate than a T-DNA from pSDM8000 in JKM115. There was no significant difference in the frequency of T-DNA transfer to these two yeast strains as was determined by T-DNA transfer experiments in which a T-DNA, that carried

15   the *KanMX* marker and the yeast 2 micron replicon, was employed. Therefore, the differences in the frequencies of T-DNA integration by HR and IR between the yeast strains YPH250 and JKM115, respectively, is most likely contributed to differences in the capacities of their HR and IR recombination machineries.

We confirmed by PCR that the T-DNA from pSDM8001 became

20   integrated at the *PDA1* locus by homologous recombination (data not shown). In order to find out whether the T-DNA from pSDM8000 had integrated randomly by IR yeast target sites for integration were determined from 8 G418 resistant YPH250 colonies by Vectorette PCR (for detailed description see materials and methods). Chromosomal DNA was isolated and digested with a

25   restriction enzyme that cuts within the T-DNA. A Vectorette was ligated to the digested DNA and a PCR was performed using a T-DNA specific and a Vectorette specific primer. The PCR product obtained was cloned into pGEMTEasy and sequenced using a T-DNA specific primer. The position of the T-DNA insertion was determined by basic BLAST search of the yeast genome

30   (http://www-genome.stanford.edu/SGD). We were thus able to map the position

22

of the T-DNA insertions of all 8 G418 resistant colonies analyzed. They were present at different positions spread out over the genome. Comparison of the T-DNA sequence and yeast target site sequences did not reveal any obvious homology. These data show that the T-DNA from pSDM8000 had integrated

5   via an IR mechanism as expected.

The following characteristics have previously been observed for T-DNAs integrated by IR: a) the 3' end of the T-DNA is usually less conserved compared to the 5' end, b) microhomology is sometimes present between the T-DNA ends and the target site, c) integration is often accompanied by small

10   deletions of the target site DNA (Matsumotot et al. 1990), (Gheysen et al. 1991, (Mayerhofer et al. 1991), (Bundock and Hooykaas 1996). Similar characteristics were seen in the currently analyzed 8 T-DNA insertions. In 3 strains we observed microhomology of 2 – 6 bp between the LB and yeast target site (figure 1, WT.51 was taken as an example). In 5 strains deletions of

15   1 – 5 bp of yeast target site DNA was found and we observed deletions, varying from 1 – 112 bp, of the 3' end of the T-DNA in 7 out of 8 analyzed strains. In only 1 strain the 3' end appeared to be intact. The 5' end of the T-DNA was conserved in almost all strains. In only 2 strains we observed small deletions of 1 and 2 bp at the 5' end of the T-DNA.

20   Thus, we can conclude that the T-DNA from pSDM8000 had integrated via the same IR mechanism described before.

Table 2: Frequencies of T-DNA integration by IR relative to integration by HR in recombination defective yeast strains

Strain	Genotype	Freq. of IR[a]	Freq. of HR	Absolute IR/HR ratio[b]	Standardized IR/HR ratio[c]
YPH250	WT	1.6×10^{-7}	2.4×10^{-5}	0.007	1
YPH250 rad51	rad51Δ	1.4×10^{-7}	1.5×10^{-6}	0.09	14
YPH250 rad52	rad52Δ	3.8×10^{-7}	2.5×10^{-6}	0.15	23
YPH250 yku70	yku70Δ	$<3.2 \times 10^{-9}$	3.3×10^{-5}	<0.0001	<0.01
YPH250 rad50	rad50Δ	8.0×10^{-9}	3.5×10^{-5}	0.0002	0.03
YPH250 lig4	lig4Δ	3.7×10^{-9}	2.3×10^{-5}	0.0002	0.02
JKM115	WT	1.2×10^{-5}	1.8×10^{-4}	0.07	1
JKM129	xrs2Δ	2.7×10^{-7}	5.1×10^{-5}	0.005	0.08
JKM138	mre11Δ	2.9×10^{-7}	7.5×10^{-5}	0.004	0.06
YSL204	sir4Δ	1.5×10^{-7}	1.8×10^{-5}	0.008	0.13

[a] Averages of 2 or more independant experiments are shown. Frequencies are depicted as the number of G418 resistant colonies devided by the output number of yeast cells (cells/ml).

[b] The frequency of T-DNA integration by IR (pSDM8000) devided by the frequency of T-DNA integration by HR (pSDM8001).

[c] The ratio of IR/HR in the mutant strain devided by the ratio of IR/HR in the wildtype strain.

24

2. Host-specific factors involved in random T-DNA integration.

The observation that the T-DNA from pSDM8000 integrates by IR into the yeast genome allowed us to use this system to study the effect of host factors on the process of integration. Many proteins involved in various forms

5 of DNA recombination have been identified in yeast. In order to determine the roles of a representative set of these enzymes in T-DNA integration, we compared T-DNA transfer and integration in wildtype yeasts with that of strains carrying disruptions of the genes encoding several recombination proteins. The *RAD51, RAD52, KU70, RAD50* and *LIG4* genes were deleted

10 from YPH250 using the one step disruption method. Yeast strains carrying deletions in *MRE11, XRS2* and *SIR4* in the JKM115 background were kindly provided by Dr. J. Haber. The results of cocultivations with these yeast strains are shown in Table 2.

In *rad51* and *rad52* mutants, which are impaired in homologous

15 recombination, the frequency of T-DNA integration by HR was 16- and 9-fold lower, respectively, than observed for the wildtype (Table 2). This showed that *RAD51* and *RAD52* play a role in T-DNA integration by homologous recombination. In the IR defective *ku70, rad50, lig4, mre11, xrs2* and *sir4* mutants the frequency of T-DNA integration by HR did not differ significantly

20 from that observed for wildtype (Table 2). This showed that these genes do not play a role in T-DNA integration by homologous recombination.

The frequency of T-DNA integration by IR in a *rad51* mutant did not differ significantly from that observed for wildtype, whereas in a *rad52* mutant the frequency was about 2-fold higher (Table 2). This showed that *RAD51* and

25 *RAD52* are not involved in T-DNA integration by IR. The product of the *RAD52* gene may compete with IR-enzymes for the T-DNA and thereby inhibits integration by IR to some extent. Strikingly, in *rad50, mre11, xrs2, lig4* and *sir4* mutants the frequency of T-DNA integration by IR was reduced dramatically: 20- to more than 40-fold (Table 2). T-DNA integration by IR

30 seemed to be completely abolished in the *ku70* mutant. We did not obtain any

G418 resistant colonies from several cocultivation experiments. This strongly suggests that KU70 plays an important role in random T-DNA integration in yeast.

5   Since T-DNA integration by HR is normal in these mutants, these results clearly show that the yeast genes *KU70, RAD50, MRE11, XRS2, LIG4* and *SIR4* are involved in T-DNA integration by illegitimate recombination.

3. Chromosomal distribution of integrated T-DNA copies in IR defective *S.cerevisiae.*

10   From several cocultivation experiments with the *rad50, mre11, xrs2, lig4* and *sir4* mutants we obtained a small number of G418 resistant colonies. The T-DNA structure was determined for a number of these lines. To this end chromosomal DNA was isolated from these G418 resistant colonies and subjected to vectorette PCR to amplify junction sequences of genomic DNA and

15   T-DNA. PCR products were cloned and sequenced. The yeast sequences linked to the T-DNA were used in a BLAST search at http://www-genome.stanford.edu/SGD to determine the T-DNA integration sites.

Strikingly, analysis of LB/genomic DNA junctions revealed that in 2 out of 3 *rad50*, 4 out of 6 *mre11* and 2 *xrs2* strains analyzed, T-DNAs had

20   integrated in telomeres or subtelomeric regions (*rad50k.1, rad50k.6, mre11k.8, mre11k.11, mre11k.14, mre11k.17, xrs2k.1* and *xrs2k.17*; Table 3 and figure 1). *S. cerevisiae* telomeres generally consist of one or more copies of the Y' element followed by telomerase-generated C(1-3)A/TG(1-3) repeats (Zakian 1996). In 2 *rad50* strains, 2 *mre11* strains and 1 *xrs2* strain the LB was found to be fused

25   to this typical telomerase-generated C(1-3)A/TG(1-3) repeat (*rad50k.1, rad50k.6, mre11k.14, mre11k.17* and *xrs2k.1*; figure 1). Besides, we also found one T-DNA insertion in a Ty LTR element in the *mre11* mutant and 2 insertions in the rDNA region, present in chromosome XII, in the *mre11* and *rad50* mutants (*mre11k.5, mre11k.4* and *rad50k.5*, respectively; Table 3 and

30   figure 1).

The 3' end of the T-DNA was truncated in all strains. Deletions of 3 – 11 bp of the 3'end of the T-DNA were observed (figure 1). Microhomology between the 3' end of the T-DNA and yeast target site was only found in 2 lines (5 bp in *mre11k.4* and 4 bp in *mre11k.14*; figure 1). For the T-DNA copies present at

5   the yeast telomeres, the RB/genomic DNA junction sequences could not be obtained from these strains using vectorette PCR. This was only possible for the *rad50* and *mre11* strains carrying the T-DNA in the rDNA region on chromosome XII. In both strains the RB was intact and no homology between the 5' end of the T-DNA and the yeast target site was found (data not shown in

10   figure 1).

Previously, target sites for T-DNA integration in the genome of *S.cerevisiae* strain RSY12 were determined (Bundock and Hooykaas 1996), (Bundock 1999. In 4 out of 44 strains analyzed, T-DNA copies were integrated in rDNA, Ty LTR elements (in 2 strains) and a subtelomeric located Y'

15   element, respectively. In addition, we determined the position of T-DNA integration in ten *S.cerevisiae* YPH250 strains. We did not find any T-DNA insertions in rDNA, LTR elements or subtelomeric/telomeric regions amongst these ten. Pooling all insertions analyzed in wildtype (54), in 2 out of 54 strains analyzed (4%) insertions were found in a Ty LTR element and in two

20   other strains in the rDNA repeat (2%) and a subtelomeric region (2%), respectively. In contrast, we report here that T-DNA in yeast strains mutated in *RAD50, MRE11* or *XRS2* T-DNA integrates preferentially in (sub)telomeric regions (8 out of 11 lines: ~73%) of *rad50, mre11* and *xrs2* mutants (table 3). From the remaining strains two T-DNAs were present in rDNA and one in a

25   Ty LTR element, respectively. Apparently, the rDNA repeat is also a preferred integration site in these mutants (~18% vs. ~2% in the wildtype).

Telomeres consist of a large array of telomerase-generated C(1-3)A/TG(1-3) repeats (~350 bp). In the subtelomeric regions two common classes of Y' elements, 6.7 and 5.2 kb, can be found (in most strains chromosome I does

30   not contain Y') (Zakian and Blanton 1988), making the average size of these

regions ~6,0 kb. Thus, the yeast genome contains (16 x 2 x 0.35) + (15 x 2 x
6,0) = 191 kb of subtelomeric/telomeric sequences. The yeast genome is 12,052
kb in size, which means that only 1.6% of the genome consists of
subtelomeric/telomeric sequences. In accordance with this, we observed in only
5   2% of the wildtype strains T-DNA copies inserted in a subtelomeric region,
which we would expect on the basis of random T-DNA integration. In contrast,
in the *rad50, mre11* and *xrs2* mutants 73% of the T-DNA insertions were found
in the (sub)telomeric region.

Analysis of 7 lines revealed that in the *sir4* mutant T-DNA was
10 integrated randomly into the yeast genome. So, although *SIR4* has an effect on
the efficiency of T-DNA integration by IR, the pattern of T-DNA distribution in
the transformants seems similar as in the wildtype strain. In the *sir4* mutant
T-DNA integration by IR was characterized by truncation of the 3' end of the
T-DNA, deletions at the target site and microhomology between the LB and
15 the target site (data not shown), like this was observed for T-DNA integration
by IR in the wildtype.

These results clearly show that in the *rad50, mre11* and *xrs2* mutants
the T-DNA, if integrated at all, becomes preferentially inserted in telomeres or
subtelomeric regions and that the genomic distribution of integrated T-DNAs
20 is altered when compared to wildtype. However, disruption of *SIR4* did affect
the efficiency of T-DNA integration by IR, but not the characteristics of this
process.

28

Table 3: genomic distribution of T-DNA integrated by IR in *rad50, mre11* and *xrs2* mutants in comparison with the wildtype after T-DNA transfer from pSDM8000

Yeast strain	(Sub)Telomeric region	LTR	rDNA	Elsewhere
rad50 mutant	2	0	1	0
mre11 mutant	4	1	1	0
xrs2 mutant	2	0	0	0
wildtype	1	2	1	50

5

·DESCRIPTION OF FIGURES

Figure 1: Junction sequences of T-DNA and *S.cerevisae* genomic DNA.
S.cerevisiae YPH250 (WT), *rad50, mre11* and *xrs2* strains were cocultivated

5   with LBA1119(pSDM8000). G418 resistant colonies were obtained.
Chromosomal DNA was isolated and subjected to Vectorette PCR to determine
the sequence of genomic DNA flanking the T-DNA. Position of T-DNA
integration was determined by basic BLAST search of the yeast genome at
http:/www.genome-stanford.edu/SGD. The Watson strand of genomic DNA

10   that is fused to the LB or RB is shown in italics. Bold sequences represent
sequence homology between the LB and target site. The filler DNA sequence is
underlined and depicted in italics. The numbers above the LB sequences
represents the number of bp deleted from the LB. Tel. = telomeric, Subtel. =
subtelomeric and Int. = intergenic.

15

Figure 2: Alignment of KU70 homologues. Sc = *Saccharomyces
cerevisiae*, Hs = *Homo sapiens* and At = *Arabidopsis thaliana*. Perfect identity
is depicted as black boxes, homology is depicted as grey boxes and dashes were
used to optimise alignment.

20

Figure 3: Alignment of LIG4 homologues. Sc = *Saccharomyces cerevisiae,*
Hs = *Homo sapiens* and At = *Arabidopsis thaliana*. Perfect identity is depicted
as black boxes, homology is depicted as grey boxes and dashes were used to
optimise alignment.

25

Figure 4: Alignment of MRE11 homologues. Sc = *Saccharomyces cerevisiae*, Hs = *Homo sapiens* and At = *Arabidopsis thaliana*. Perfect identity is depicted as black boxes, homology is depicted as grey boxes and dashes were used to optimise alignment.

5

Figure 5: Alignment of RAD50 homologues. Sc = *Saccharomyces cerevisiae*, Hs = *Homo sapiens* and At = *Arabidopsis thaliana*. Perfect identity is depicted as black boxes, homology is depicted as grey boxes and dashes were used to optimise alignment.

10

Figure 6: Alignment of XRCC4 homologues. Sc = *Saccharomyces cerevisiae*, Hs = *Homo sapiens* and At = *Arabidopsis thaliana*.

15

REFERENCE LIST

Alani, E., Subbiah, S., and Kleckner, N. 1989. The yeast *RAD50* gene encodes a predicted 153-kD protein containing a purine nucleotide-binding domain and two large heptad-repeat regions. Genetics 122:47-57.

Bechtold, N., Ellis, J. & Pelletier, G. (1993). *In planta Agrobacterium* mediated gene transfer by infiltration of adult *Arabidopsis thaliana* plants. C R Acad Sci Paris Life Sci *316*, 1194-1199.

Bundock, P. 1999. *Agrobacterium tumefaciens*-mediated transformation of yeast and fungi. Thesis Leiden University.

Bundock, P., den Dulk-Ras, A., Beijersbergen, A., and Hooykaas, P. J. J. 1995. Trans-kingdom T-DNA transfer from *Agrobacterium tumefaciens* to *Saccharomyces cerevisiae*. EMBO J. 14:3206-3214.

Bundock, P. and Hooykaas, P. J. J. 1996. Integration of *Agrobacterium tumefaciens* T-DNA in the *Saccharomyces cerevisiae* genome by illegitimate recombination. Proc. Natl. Acad. Sci U.S.A 93:15272-15275.

Capecchi, M.R. (1980) High efficiency transformation by direct microinjection of DNA into cultured mammalian cells. Cell 22, 479-488.

de Groot, M. J. A., Bundock, P., Hooykaas, P. J. J., and Beijersbergen, A. G. M. 1998. *Agrobacterium tumefaciens*-mediated transformation of filamentous fungi. Nat. Biotech. 16:839-842.

den Dulk-Ras, A. and Hooykaas, P. J. J. 1995. Electroporation of *Agrobacterium tumefaciens*. Methods Mol. Biol. 55:63-72.

32

D'Halluin, K., Bonne, E., Bossut, M., Beuckeleer, M.D., and Leemans, J. (1992). Transgenic maize plants by tissue electroporation. Plant Cell *4*, 1495-1505.

5   Dunwell, J.M. (1999). Transformation of maize using silicon carbide whiskers. Meth. Mol. Biol. 111: 371-382.

Gheysen, G., Villarroel, R., and Van Montagu, M. 1991. Illegitimate recombination in plants: a model for T-DNA integration. Genes Dev. 5:287-
10   297.

Gietz, D., St.Jean, A., Woods, R. A., and Schiestl, R. H. 1992. Improved method for high efficiency transformation of intact yeast cells. Nucleic Acids Res. 20:1425.

15

Gietz, R.D. & Woods, R.A. (2001) Genetic transformation of yeast. Biotechniques *30*, 816-820.

Hanover, J.A. (1992) The nuclear pore: at the crossroads. FASEB J. 6, 2288-
20   2295

Hansen, G. and Chilton, M. D. 1999. Lessons in gene transfer to plants by a gifted microbe. Curr. Top. Microbiol. Immunol. 240:21-57.

25   Hartung, F. and Puchta, H. 1999. Isolation of the complete cDNA of the *MRE11* homologue of *Arabidopsis* (Accession No. AJ243822) indicates conservation of DNA recombination mechanisms between plants and other eukaryotes. Plant Phys. 121:312-312.

Hooper, M.L. (19... Embryonal stem cells: introducing planned changes into the animal germline, in *Modern genetics* Vol.1, 1st ed., Evas, H.J., Chur: Harwood Academic Publishers.

5   Hooykaas, P. J. J. and Beijersbergen, A. G. M. 1994. The virulence system of *Agrobacterium tumefaciens*. Annu. Rev. of Phytopathol. 32:157-179.

Jones, J. S. and Prakash, L. 1990. Yeast *Saccharomyces cerevisiae* selectable markers in pUC18 polylinkers. Yeast 6:363-366.

10

Junop MS, Modesti M, Guarne A, Ghirlando R, Gellert M, Yang W. (2000) Crystal structure of the Xrcc4 DNA repair protein and implications for endjoining. EMBO J.19, 5962-5970.

15   Kaeppler, H.E., Gu, W., Somers, D.A., Rines, H.W., Cockburn, A.F. (1990) Silicon carbide fiber-mediated DNA delivery into plant cells. Plant Cell Rep. 9, 415-418

Kegel A, Sjostrand JO, Astrom SU. (2001) Nej1p, a cell type-specific regulator
20   of non-homologous end joining in yeast. Curr Biol.11:1611-1617.

Klein, T.M., Arentzen, R., Lewis, P.A., and Fitzpatrick-McElligott, S. (1992). Transformation of microbes, plants and animals by particle bombardment. Biotechn. *10*, 286-291.

25

Kunik, T., Tzfira, T., Kapulnik, Y., Gafni, Y., Dingwall, C., and Citovsky, V. (2001). Genetic transformation of HeLa cells by *Agrobacterium*. Proc.Natl.Acad.Sci.U.S.A. *98*, 1871-1876.

Lee, S. E., Paques, F., Sylvan, J., and Haber, J. E. 1999. Role of yeast *SIR* genes and mating type in directing DNA double-strand breaks to homologous and non-homologous repair paths. Curr. Biol. 9:767-770.

5   Lin, C.H., Xiao, L., Hou, B.H., Ha, S.B., and Saunders, J.A. (1997). Optimization of electroporation conditions for expression of GUS activity in electroporated protoplasts and intact plant cells. Plant Physiology and Biochemistry, *35*, 959-968.

10   Marangoni, E., Foray, N., O'Driscoll, M., Douc-Rasy, S., Bernier, J., Bourhis, J., Jeggo, P. (2000) A KU80 fragment with dominant negative activity imparts a radiosensitive phenotype to CHO-K1 cells. Nucl. Acids. Res. 28, 4778-4782.

       Matsumoto, S., Ito, Y., Hosoi, T., Takahashi, Y., and Machida, Y. 1990.
15   Integration of Agrobacterium T-DNA into a tobacco chromosome: possible involvement of DNA homology between T-DNA and plant DNA. Mol. Gen. Genet. 224:309-316.

       Mayerhofer, R., Koncz-Kalman, Z., Nawrath, C., Bakkeren, G., Crameri, A.,
20   Angelis, K., Redei, G. P., Schell, J., Hohn, B., and Koncz, C. 1991. T-DNA integration: a mode of illegitimate recombination in plants. EMBO J. 10:687-704.

       Milne GT, Jin S, Shannon KB, Weaver DT (1996) Mutations in two Ku
25   homologs define a DNA end-joining repair pathway in *Saccharomyces cerevisiae*. Mol Cell Biol. 8, 4189-4198.

       Moore, J. K. and Haber, J. E. 1996. Cell cycle and genetic requirements of two pathways of nonhomologous end-joining repair of double-strand breaks in
30   Saccharomyces cerevisiae. Mol. Cell. Biol. 16:2164-2173.

Offringa, R. 1992. Gene targeting in plants using the *Agrobacterium* vector system. Thesis Leiden University.

Offringa, R., de Groot, M. J., Haagsman, H. J., Does, M. P., van den Elzen, P. J., and Hooykaas, P. J. J. 1990. Extrachromosomal homologous recombination and gene targeting in plant cells after *Agrobacterium* mediated transformation. EMBO J. 9:3077-3084.

Ozeki, K., Kyoya, F., Hizume, K., Kanda, A., Hamachi, M. & Nunokawa, Y. (1994) Transformation of intact *Aspergillus niger* by electroporation. Biosci Biotechnol Biochem *12*, 2224-2227.

Paull TT, Gellert M. (1999) Nbs1 potentiates ATP-driven DNA unwinding and endonuclease cleavage by theMre11/Rad50 complex. Genes Dev. 13, 1276-1288.

Picard, D., Salser, S.J., Yamamoto, K.R. (1988) A movable and regulatable inactivation function within the steroid binding domain of the glucocorticoid receptor. Cell 54, 1073-1080.

Raikhel, N. (1992) Nuclear targeting in plants. Plant physiol. 100, 1627-1632.

Rothstein, R. 1991. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol. 194:281-301.

Schiestl, R.H., Manivasakam, P., Woods R.A. and Gietz, R.D. 1993. Introducing DNA into yeast by transformation. Methods: A companion to methods in enzymology 5: 79-85.

36

Schiestl, R. H., Zhu, J., and Petes, T. D. 1994. Effect of mutations in genes affecting homologous recombination on restriction enzyme-mediated and illegitimate recombination in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 14:4493-4500.

5

Schild, D., I. L. Calderon, C. R. Contopoulou, and R. K. Mortimer. 1983. Cloning of yeast recombination repair genes and evidence that several are nonessential genes, p. 417-427. *In*: E. C. Friedberg and B. A. Bridges (eds.), Cellular responses to DNA damage. Alan R. Liss Inc., New York.

10

Schnorf, M., Neuhaus-Url, G., Galli, A., Iida, S., and Potrykus, I. (1991). An improved approach for transformation of plant cells by microinjection: molecular and genetic analysis. Transgenic Res. *1*, 23-30.

15 Sikorski, R. S. and Hieter, P. 1989. A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae*. Genetics 122:19-27.

Steensma, H. Y., Holterman, L., Dekker, I., van Sluis, C. A., and Wenzel, T. J.
20 1990. Molecular cloning of the gene for the E1 alpha subunit of the pyruvate dehydrogenase complex from *Saccharomyces cerevisiae*. Eur. J. Biochem. 191:769-774.

Taccioli GE, Gottlieb TM, Blunt T, Priestley A, Demengeot J, Mizuta
25 R,Lehmann AR, Alt FW, Jackson SP, Jeggo PA (1994) Ku80: product of the XRCC5 gene and its role in DNA repair and V(D)J recombination. Science. 265, 1442-1445.

Templeton, N.S., Roberts, D.D. & Safer, B. (1997) Efficient gene targeting in
30 mouse embryonic stem cells. Gene Ther. *4*, 700-709.

Teo, S.-H. and Jackson, S.P. (2000) Lif1p targets the DNA ligase Lig4p to sites of DNA double-strand breaks. Curr. Biol. 10, 165-168

Valencia, M., Bentele, M., Vaze, M.B., Herrmann, G., Kraus, E., Lee, S.E., Schar, P., Haber, J.E. (2001) Nature 414, 666-669

Vergunst, A.C., De Waal, E.C., and Hooykaas, P.J.J. (1998). Root transformation by *Agrobacterium tumefaciens*. In Arabidopsis protocols. J. Martinez-Zapater and J. Salinas, eds. (Totowa, NJ: Humana Press Inc),

Vergunst, A. C., Schrammeijer, B., den Dulk-Ras, A., de Vlaam, C. M. T., Regensburg-Tuïnk, T. J. G., and Hooykaas, P. J. J. 2000, VirB/D4 dependent protein translocation from *Agrobacterium* into plant cells. Science: in press.

Wach, A., Brachat, A., Pohlmann, R., and Philippsen, P. 1994. New heterologous modules for classical or PCR-based gene disruptions in *Saccharomyces cerevisiae*. Yeast 10:1793-1808.

WO 00/68404, Pioneer Hi-bred.

WO 00/12716, Pioneer Hi-bred.

Zakian, V. A. 1996. Structure, function, and replication of *Saccharomyces cerevisiae* telomeres. Annu. Rev. Genet. 30:141-172.

Zakian, V. A. and Blanton, H. M. 1988. Distribution of telomere-associated sequences on natural chromosomes in *Saccharomyces cerevisiae*. Mol. Cell. Biol. 8:2257-2260.

Zrenner, R., Willmitzer, L., and Sonnewald, U. 1993. Analysis of the
expression of potato uridinediphosphate-glucose pyrophosphorylase and its
inhibition by antisense RNA. Planta, 190:247-252

5

Claims

1.   A method to direct integration of a nucleic acid of interest to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in a eukaryote with a preference for non-homologous recombination, comprising steering an integration pathway towards homologous recombination.

2.   A method to direct nucleic acid integration according to claim 1, comprising providing a mutant of a component involved in non-homologous recombination.

3.   A method to direct nucleic acid integration according to claim 1 or 2, comprising inhibiting a component involved in non-homologous recombination.

4.   A method according to claim 2 or 3 wherein said component involved in non-homologous recombination comprises *ku70, rad50, mre11, xrs2, lig4* or *sir4*.

5.   A method to direct integration of a nucleic acid of interest to a pre-determined site according to anyone of claims 1 to 3, wherein said nucleic acid sequence of interest is essentially replacing a sequence within said eukaryote.

6.   A method to direct integration of a nucleic acid of interest to a pre-determined site according to claim 5, wherein said component involved in non-homolgous recombination comprises *rad50* or *xrs2*.

7.   A method to direct integration of a nucleic acid of interest to a sub-telomeric and/or telomeric region in a eukaryote with a preference for non-homologous recombination by providing a mutant of a component involved in non-homologous recombination.

8.   A method to direct integration of a nucleic acid of interest to a sub-telomeric and/or telomeric region in a eukaryote with a preference for non-homologous recombination, comprising inhibiting a component involved in non-homologous recombination

9.   A method to direct integration according to claim 7 or 8 wherein said component involved in non-homologous recombination comprises *rad50*, *mre11* or *xrs2*.

10.   A method according to anyone of claims 1 to 9 wherein said eukaryote comprises yeast, a fungus or an animal.

11.   A method according to anyone of claims 1 to 10, wherein said nucleic acid of interest is delivered to a cell of said eukaryote by *Agrobacterium*.

12.   A method according to anyone of claims 1-11 comprising transiently inhibiting integration via non-homologous recombination.

13.   A method according claim 12 wherein said transiently inhibiting is provided by an *Agrobacterium* Vir-fusion protein capable of inhibiting a component involved in non-homologous recombination.

14.   A method to direct nucleic acid integration according to claim 13 wherein said *Agrobacterium* Vir fusion protein comprises VirF or VirE2.

14.   A method according to claim 13 or 14 wherein said component involved in non-homologous recombination comprises *ku70*, *rad50*, *mre11*, *xrs2*, *lig4* or *sir4*.

15.   A method according to anyone of the aforegoing claims wherein said nucleic acid of interest comprises an inactive gene to replace an active gene.

16.   A method according to anyone of claims 1-14, wherein said nucleic acid of interest comprises an active gene to replace an inactive gene.

17.   A method according to anyone of claims 1-14, wherein said nucleic acid of interest encodes a therapeutic proteinaceous substance.

18.   A method according to anyone of claims 1-14, wherein said nucleic acid of interest encodes a substance conferring resistance for an antibiotic substance to a cell.

19.   A method according to anyone of claims 1-14, wherein said nucleic acid of interest confers a desired property to said eukaryotic cell.

20.   A method according to anyone of the aforegoing claims wherein said nucleic acid of interest is part of a gene delivery vehicle.

21.   Use of a method according to anyone of claims 1 to 20 for improvement of gene targeting efficiency.

Fig. 1

| | LB'   T-DNA   RB' | |
Strain	CAGGATATATTCAATTGTAAAT-CTC---CGA-GG	Chromosome, coordinate and location
WT.51	-4 5' *ATTGTATTATATATTCAATTGTAAAT-CTC---CGA-GGTA* 3'	XIV, 185311 (1 bp of target site DNA deleted), int. region
rad50k.1	-6 5'*TGTGGGTGTGATATTCAATTGTAAAT-CTC---CGA-GG* 3'	XV, 1091277, tel. region
rad50k.5	-7 5' *GGGGGCATCAGTATTCAATTGTAAAT-CTC---CGA-GG* 3'	XII, 465986, rDNA region
rad50k.6	-25 5' *GAGGTAGATGTGAGAGAGTGTGTGTGGGTGTGAAGTCGA* 3'	XV, 1091276, tel. region
mre11k.4	-3 5' *TCTGGTAGATATATTCAATTGTAAAT-CTC---CGA-GG* 3'	XII, 459692/468829, rDNA region
mre11k.5	-8 5' *CACATATTTCTCATTCAATTGTAAAT-CTC---CGA-GG* 3'	VII/X/XIII, 536090 OR 541678/ 472487 OR 483659/196667, LTR
mre11k.8	-11 5' *CGACTACTTTATAT<u>CC</u>AATTGTAAAT-CTC---CGA-GG* 3'	XIV, 6060, subtel. region
mre11k.11	-7 5' *GAAGAACCCATTATTCAATTGTAAAT-CTC---CGA-GG* 3'	XIV, 4866, subtel. region
mre11k.14	-7 5' *TGGGTGTGGG**TTATT**CAATTGTAAAT-CTC---CGA-GG* 3'	VIII, 562588, tel. region
mre11k.17	-9 5' *TGGGTGTGGTGTGTTCAATTGTAAAT-CTC---CGA-GG* 3'	XII, 5727, subtel. region
xrs2k.1	-10 5' *TGTGTGGGTGTGGGTCAATTGTAAAT-CTC---CGA-GG* 3'	IX/X, 69/52, tel. region
xrs2k.17	-1 5' *CGTCAAGGATATATTCAATTGTAAAT-CTC---CGA-GG* 3'	XII, 1071797, subtel. region

Fig. 2

```
Sc     1 ------MRSVTNAFGNSGELN   E GYRK DI G  FC LSETMFKESS LEYKS LLEI ES DE MSQL I RP
Hs     1 MSGWESYYKTEGDEEAEE QE    A GDYK SG S FL AS KAM ESQS DELT PFDMS QC QS YISK I SD
At     1 -----------------------------------------------------------------------------------


Sc    75 GTA  CYF YCN  DAKEGIYE  P R NATFMK NDLLEDLSS RISLYD FMFQQTGSEKQ RLSVL TF LDTF L
Hs    80 RDL  VV YGTE  KNSVNFKN  V Q DNPGA    EL QFK QQGQKR QDMMGH GS YS  SEVLW--VCANL F
At     1 -------------------------------------------------------------- NSI-YS LW--VAQAL L


Sc   155 E  PGQKQ SNKR FLFT ID PQEAQD-IDERA LRR---LTI DL F DNK NFATFF GYADKPFDN-EFYSDI QL SH
Hs   155 S   --QFK SHKR MLFTNEDNPHGND--SA AS ART--- AGDLRDTGI F L H KK GG-FDI LFY DIIS E-
At    16 RKG--SLKT DKR FLFTNEDDPFGSMRISV EDMTRTTLQ AKDAQDLGIS L L PLSQPDKQF I LFY D GL NS-


Sc   230 T ENTGLD EFDGPS KP AKY KS L KKEV RI FQCPL LD KTNF GV KGYT YT EKA GVRYK  EHE IR
Hs   226 -DE----DLRVHFEE SKLED-- LR AKETRKRA SR KLKLNKD-- ISVGIYNL Q A KPP----P KLYR TN
At    93 -DE----L EFMPSVGQKLED-- DQ  VLA RIAKR TF  CG--- SIE NGYAL RPA PGS--- T DSTTN


Sc   310 QEAY KR  NPITG-ED TGKTV YPYGDLDINL DS DQIVMEAYTQKD FL   GF SSSKS Y NN DKS F V
Hs   294 E VK KT TFNTSTGGL PSD KRSQIYGSR T LEKEETEE KRFD---DPGLML GFKPLV-LLKKH HY PSL Y
At   161 L VKVERS CTDTG-A QDPIQR QPYKNQ I F VEELSQVKRIS---T HL L GFKPLS-CLKD HNL PS Y


Sc   389 P AKYEGSIRTLA LL I RKK IA L GKL SN H SLYTLSPSS-- KDYN-----EGFYLYR PFLD IRKFPS L
Hs   370 P ESL VIGS TLF SALL IKC EKEVAA CRYTP RNI PPYFVALVPQ EE DQKIQVTPPGF L LPFADDKRK ---
At   236 PS KE VIGS RAFIALH S  QL FA A YG--GT PPRLVALVAQ E SDGGQVEPPGI  LP A DIRD DEL


Sc   462 LSYDDG E K DYDNMKKVTQ IMGYFNL DGYNPSD KNP LQ HY LHDYL ---  ET TFDE NET PNTKK R-- 
Hs   447 ----P  E I A P Q GKMKAI E RFT--YRSDS ENP LQQH NLEA ALDLM PEQAVDLT PKVE MN -R G
At   313 HSK- GVAXPRA DIQ KKASA MR IEL D- SVCQFANPALQ HYA LQA ALDENP RETRDETLP DEE MNRPA


Sc   537  EDDSL K YY IR KILES -------KS PDPIIQ LNKYVK WN--------MFYKKFN---------DDN S KEE K
Hs   520 SL DE KE  YPPDYNPEGK-VT RRHDNEGSGSKRP- V YS EEELKTHISK GT GKFTVPMLKEACR YG KSGL K Q
At   391  A  QFKQS YGDDPDEES SGAKE KS KRKAG DADDGKY YIEL-AKT----GK RDLTVVELKTYLTANN L SGKKE


Sc   593 PFDK PKFN  --
Hs   598 E  EA TK FQD
At   466 V  N  LT I GK
```

Fig. 3

```
Sc     1  MISALDSPPEPQNFAPSPDFKWCEELFVKIHEVQINGTAGGKSSFKYYEIISNVEMWRKTVGNNIYPALVLALPYR
Hs     1  --------------MTEEIKFS--------LVSLFNWIQKSKTSSQKRKFRFLDTYCKPSDFVAVR-----LIPQLY-----R
At     1  --------MTEEIKFS--------LVSLFNWIQKSKTSSQKRKFRFLDTYCKPSDFVAVR-----LIPSL-----R

Sc    81  RRIYNIKYVLIFTCSYLKLPSATEQRLKIW--QRVGKGGNLS--SLVEEAKRRAEPSSKTIDVNHYLDS
Hs    11  ERMAYGIKEVLALYIILLNLPRDKDAALLNKR--TPTGTHGAGDFIAYFVLKPR-CQKGSLTIQVNDLLDS
At    57  ERGSYGKEVLATCIALGRSRDPDARLLNWRKGGTAKAGNAGNFSLIAAEVLQRRQGASGLTIKENDLLDR

Sc   157  LGGDRFASGRFKLVKSKPFHCEMSFEKFFDILKRVGGQEKLLCHPDAQDYLSVISDLKVTSKLYD
Hs    88  ASN---NSAKRKDLKKS-LLQLTSSALEQKWIRIKDLKLGSQIFSVPHNDAALLHNVTDLEKVCRQLED
At   137  LASS---ENREKLVLS-LQKTN---AQEKWIRIILKDLKLGSEIBEPHPDAEDLFNVTCDLKVCEKLED

Sc   237  PVVRLKDDDSIKVGFAFAPQLAKNLSYEKICTLHDDFEEKDGERIQHYMNYGSIKFSRRGDYTYLGAS
Hs   164  PSVGLSD--SITLFSASPMLAA-IDDEHIEKDHQSBYETKDGERMQHKD--GVYKYFSRNGYNYTTQGAS
At   210  RHQRHKR--QDIEVGKAVPQLAMFIDNAAWKGKDVAECKFDGRIQHKN--GTDIYFSRNFDDHYAHAM

Sc   317  LNSGE--QHLRFTDSKECLDGEMTFDAKRRVIPFGLVKGSAKEALSFNSNVDFHPLYVFDLYNGTSITP
Hs   239  PEEGSPFIHNAFKADLQICILDGEMAZPNTTFEQKGKFFIK---R---MEEDSDLQTCYCVFDVIMVNNKKLGH
At   286  SDLIVQN------ILVD--KCILDGEMVEDTSLRFAEFGNQIIANAAR--EGLDSHKQLCYAFDVLYVGDTSIH

Sc   395  LPLQORYLNSIPLKNIEIVRS----------------RCYVESKSLEVAISLGSEGVLKYYNSSYNVSR
Hs   313  ELLKRYBLSSIFPLPGREIVQK----------------QAHTKNVIDALNEAIKRIEGIRQPLSIYKPDKR
At   355  QLLERELIKKCKPLKGRLEVPEGGLNVHRPSGEPSWSIVVHAALVEFFKETLNRLEGIVLKDLESKEPSDR

Sc   459  NNNWKKPEYEEFGNLDIIGRDSGKKDSFMLGLLVLDEEEYKKHQGDSSEIVDHSQEKHIQNSRRVKKISFC
Hs   377  GEGWKKPEYSGLMLDLIGGYGKGS----------------RGGMSHFLCAAEKPPPGEPSV--FRLS
At   435  SGKWKKPEYR-AGALDLIIGGYGSGR----------------RGGEAQFLVALAAEANVYPRR--FRSFC

Sc   539  SNGSQEFKDRKTFGHWK-TSEVAPPASILEFG---SKIPAEWID-PSESILEIKSRSLDNTETNMQKYADNC
Hs   438  RVGGCMKELYGLKLAKYWKP-FHRKAPPSSILCGT----EPVLI-PCNSIIQIK----AAEIGPSLYKTGC
At   495  RVGGSDDELNTVSKLPYNEHPRKAPPSFYQVTNHSKEPLVWIDSPEKSIILSITS---DIRTLRSVAPY

Sc   614  TLYGGYCKIRYDKEWTCYTLDLYESRTVKSNPYQAERSQLG------LIRKKRKLISDSFHQRKQLPSNIFE
Hs   508  TLRFPRIKIRDDKEWHECTLDDLELRGKASGKLASKLYIGGDDEPQEKKRKAAPKRKVIGTIEHLKAPNLNNK
At   572  LRFPRIEKRYDKPWHECDYAFVELVNSSNGTQKQESESTQDNPKVNKSSKRGEKKNVSLPSFIQTDSIKE

Sc   688  GLLFVLSVTEDTGIRIALLEKTIVEHGGKYNILKRHSIGDVLISCKTECKALIDRG---YDHPNWVLD
Hs   588  ININIFEDFCVSGTDSQPKPLENRIABFGGYEQNPG------PDTYCIAGSENTVKNIILSNKDVPAWILE
At   652  KSIFSNIFYFNVPRSHLETFHKMVENGGKFSMNNN-----SVTCIAAESGISYQAAKRQ--DVHFSWVLD

Sc   766  CIAYKLLEPNYCFNQKMRAVAERDCLGDSENDSETKLSSKSQLSLPPMGELEIDSEFPLFLFSNR-
Hs   662  CFKTKSFPWOPMHCPSTELFAYDCGDSYIDTDINQLKEVSGIKSNEQTPEEASLADLEYRYSWDCS
At   725  CCSRNPELPKYFHDARTLQDIDELSDSYWDDLEGLKQVLSNAKSS--EDSKSIDYYKKLCPEKRWSC-

Sc   845  ---AYVPRRKSEDDIIFKIKLFGGTDQQSLCNIIIPYDPLI---DCNEHEKQIKASDTI------
Hs   742  PLSFRRHTVDYAVINISTKNEGTAIKALERFHGKVVSCLAEGVSHGEDISADFKAFRTF------
At   802  ---LSCCVYYPYSQTSTEEEALLGIMAKRLMLELAGKVNNLAASHLAAEPLDFTLVSFSEMEKR

Sc   913  ---PKIAVVAPEWVHSIENCPEEDFPNY-------------------------------
Hs   816  ---KFKESWVTSIDCEQEEQY-----------------------------------
At   878  LLLKRLVVSSHWSLEEKLCEDVYTRPKYMEESDTEESDKSEHDTTEVASQGSAQTKEPASSKIAITSSRGR
```

Fig. 3, Contd.

```
Sc        -----------------------------------------------------------------
Hs        -----------------------------------------------------------------
At   957  SNTRAVKRGRSSTNSLQRVQRRRGKQPSKISGDETEESDASEEKVSTRLSDIAEETDSFGEAQRNSSRGKCAKRGKSRVG


Sc        -----------------------------------------------------------------
Hs        -----------------------------------------------------------------
At  1037  QTQRVQRSRRGKKAAKIGGDESDENDELDGNNNVSADAEEGNAAGRSVENEETREPDIAKYTESQQRDNTVAVEEALQDS


Sc        -----------------------------------------------------------------
Hs        -----------------------------------------------------------------
At  1117  RNAKTEMDMKEKLQIHEDPLQAMLMKMFPIPSQKTTETSNRTTGEYRKANVSGECESSEKRKLDAETDNTSVNAGAESDV

Sc        -----------------------
Hs        -----------------------
At  1197  VPPLVKKKKVSYRDVAGELLKDW
```

Fig. 4

```
Sc     1  MDYPDP-D---TURILTTDNHUGYNENDPITGDDSUKTEHEUMLAKNNUVDMUQSGDLFHVNKPSUKULYQVLKTLR
Hs     1  MSTADALDDENTUILVATDIHLGUMEKDAARGNDUFVTLEIURLAQENEVDFULLGGDLFHENKPSRKTLHTCLEULR
At     1  MSREDFSD---TURULVATDCHLGYMEKDEIRRHDSFKAFUEICSUAEEKUVDFULLGGDLFHENKPSRTTLVKAUEULR

Sc    77  LCCMGDKPCULEULSDPSQVFHUDEFTUVNYEDPNFNISIPVEGISGNHDDAUGDSLLCPUDILHATGLUNHFGKUE--
Hs    81  UYCMGDUPVQFEULSDQSVNFGUSKFPWVNYUDGNUNISIPVFSIHGNHDDPUGADALCAUDILSCAGUVNHFGRSUS--
At    78  UHCUNDKPVQFUUSDQUVNFQN-AUGUVNYEDPHFNUGUPVFSIHGNHDDPAGVDNLSAUDILSACNLVNYUEGKULGG

Sc   155  --SUKIKUVPULUQKGSTKUALYGLUAURDERLURTUKD-GUVTUEVPTURE----GUWFNLUCUHQNUUGHTNTAUFUPE
Hs   159  --VUKIDUSPULUQKGSTKUALYGLGSIUDERLURMFVN-UKVTMURPKEDE----NSWFNLUVUHQNRUKHGSTNFIPE
At   157  SGVGQITUYPUULUKKGSTTUALYGLGNIRDERLNRMUQTPUVQUURPEUQEGCDVSUWENUUVUHQNRVKSNPKNAUSE

Sc   228  QFLPDFLDUVIWGHEHECUPNLVUNPIKNUDULQPGSSVATSICUUEAQPKYVFULUIKYGEAPUMTPIPLETURTUKMK
Hs   232  QFLDDFUDUVIWGHEHECKIAPTUNEQQLUYIUQPGSSVUTSISPGEAVKKHVGLLRIK-GRKMNUHKIPLUTVRQUFUE
At   237  HFLPRFLDFUUWGHEHECUIDPQEVSGMGUHIUQPGSSVATSLIUGEUKPKHVULLUIK-GNQYUPTKIPLTUVRPUFEYT

Sc   308  SUSLQDVPHLU-RUHD---KDATSKULUEQVEEMIRDANEEUUKOUULUDDGEGDMVAEUPKPLURLRVDYSUPSNTQSPIDU
Hs   311  UIVLAUHPDIFNPDUPKVTUAIQSUCULEKUEEMUENA---EUEUULUNSH------UPEKPLVRLRVDYSGU--------U
At   316  UIVLKDESDI-DPND---QUUILEHLDKVUVRNUIEKA---UKAUNRS------EUKLPLVRUUVDYSG--------U

Sc   384  QVENPRRFSUUFVGRVANGUUUUQFUKURSPVTURSKKSGINGUSUSDRDVEKLFSUSCGGELEVUTLVN----DLLNKUUQL
Hs   374  EPFSVLRFSQKFVDRVANPKDIUHFUUREQEUUGG-EEINFGKUUTU------PSEGTTURVEDLVKQYFQTAUBKNUUQL
At   373  MTINPQRFGQKUVGUVANPUDTUUIFSKASK-UGRSE-ANIUDUERUU-R------PEELNQQNUEALV------AESNUKU

Sc   460  SLLPEVGLNEAVUKFVDKDEKTAUKEFUSHEUSNEVGILSTUEEFLUTDDAUEU--KALUIKQVURANSVURPTP--PUEND
Hs   447  SLIUERGUGEAVQEFVDKUBKDAUEELVKYUUE------KTU-UFLKURHIUAUU-EDKUDEEVURFRETRQUKN--TNDED
At   438  EULPVNDLDVAUUNFVUKDURLAFYSCVQYULQ------ETRGULAKUUSDAKKFEEDDLILKVGECLEERLUDRSTUPTG

Sc   536  UUU-FAFNGNGLDSFURSSNRUVRUU-SPDITQUHUDNESRUTUUISQUUESSKPTURPU-----RVR--------UAUUKKIP
Hs   517  EU--UREUUTRARALURSQSEUSAUUFSAD-DLMSUDLAEQUANDSDDUSISAAUTNKGRUUU--RGRRGUGRGQNSAUUGGUU
At   512  SUUFUSTUUTSENLTUGSSUGIANASUFSDDEDTUQUUSGLAPPTUGRRUSSUANTUUGRUUUPTRGR--GRGKASUAMKOUUT
```

```
Sc   602  -AFUSDSTVISU-DAENELUDNNDAQUDUUIDENDUIM---VSTUEEDU-ASUGLLNGRUTKUUKURPUASTUUU---UAUURGUUGR
Hs   591  -QRGRUFUSUURQQPSRNVTUKNUSUVUUDESDVUEDIFPUUISKTD-QRUSSTSSSKUIMUUQSQVSUKGVDFEUUEDDDDDP
At   590  LDSSLUFUUQSU-QRSASAUAUUAAUUKSASTUIGEDDUUS---PSUUEVUPEDUNKPDSSSEDDESTKUKGRKRPAUURGUUGR
```
```
Sc   674  USRUPP----KTUI---UGSUULAKURR--K-----------------------------
Hs   669  FMNUTSS-LRUUNRR---UIYLLALUUN-MQDTG-KMUUCYKU--------RVUY-SLRF
At   666  USGUTSKRGRKNUUSSSSUNRLLUUSKDDDEDEDDEDUUEKKUINKSQPRVTUNYUGALRR
```

Fig. 5

```
Sc    1 ------MSA YK SIQGIRSFDSNDE E---TIEFG PLTLIVG NG GKTTIIECLKYATGDLPPN -KGGVFIHDPK T
Hs    1 MLIFSVRDMFAKMSILG RSFGIEDD KQ ITFFSPLT VGP NGAGKTTIIECLKYICTGDE PPG -KGN F HDPKVA
At    1 ------MST DKM IKGIRSFDPE RN-- TFF PLTLIVG ANGAGKTTIIECLK VSCTG LPPNARSGH FIHDPKVA

Sc   72 GEKD RAQ KLAF T ANGLN VTRN QLLMKK TTTFKTLEGQ VAINNS-G S LS RSLE DA PLYLGVP KAIL
Hs   80 QETD RAQI LQF DVNGEL AVQRS VCTQKSKKTEFKTLEGV TRT-KH -GEKV LS CAE DRE ISSLGVSKA L
At   73 GET T AQIKLRF AAGKD CIRSF QLTQKA KME KA ES VLQTINPHTGEKVCLSY RCA DRE PAL GVSKAIL

Sc  151 EYVIFCHQEDSLWPLSEPSNLKKKFDEIFQAM TKAL NLKS KDMSVF IKLLRQS EHLKLDKD KA L IHQLQ
Hs  158 NNVIFCHQEDSNWPLSEGKALKQKFDEIFSATRY IKALETL Q QTQ QK EE Q ELKYLRQYK ACE R ITSK
At  153 ENVIFVHQ SNWPLQ PSTLKKKFD IFSATRYTKALEV RKL KDQ QO IKT K KLENLQTLKDAAYK RESIAQDQ

Sc  231 T QYN EVSE E QLNE TE SDK F SNQDFQK LSK VENLKNTKL S-DQV RLSNSIL LCL KP LFNLLANF
Hs  238 AQ TSSKE IVKSY ENELDP KN F E NLSK MKLDNE KAL DSRKKQ EKD SELEEK EK FQGTDEQL LYHNH
At  233 E T SSKVQ LE ES S K DAE NK MMLKD RKLQDQVSIKTAE S FK QQ QYAA P EN DTIE LKI WKSK

Sc  310 S VLM KNNQLRD ET SSLKDRQ SSL SLSNS L RRQGEL AGK TYEKN-RN LS KEA QHKFQGLSNIENSDMA
Hs  318 Q T RE KERRLVDCHRE EKLNKESRLL EK EL EQG LQLQA RHQEHIRARDSLIQS ATQLELDGFERGPFSER
At  313 EERLALLGT RK ERE VDTETTISSLH AK NY EI S LQTEAEAHMLLKNERDS IQN FHYNLGNVPSTPFSTE

Sc  389 QVNHEMSQFKAF SNDLTDT DQFAKDIQLKETNLSDLIKSITVDSQNLEY-NKKD SK HDS---EELAER LKS KSLS
Hs  398 Q KNFHKL ERQ EGEAKTA QLMND FAEKETLKQKQIDE IRDKK GL R-IIEL SE SKKQNELKN RY LQQLEG
At  393 VVLNLTNR KSR GELEMDL DKKKS ETALS AWDCY DANDRWK IE QKRAKDEIK GISKRIEEKE ER SFEFEI

Sc  466 QDS NHELE NLKTYKEKLQSWESENI PKLNQKIEEK NE IILENQ ERF DR MKTNQQA LYAKLG IKKSINTKL
Hs  476 S DR LELD ELIKAER ELS AEK SN ETLKME ISLQNEKADL FT RLDQE QO NHHTTTRTQ E TKDRADKD
At  473 S VD KQTDERE KQVQVELE KTK NSERGFESKIEQKQHE YSLE KI TL RER V AGDA DR-LLTRIDECKDRIR

Sc  546 ELQK TE LQ DSRIRQ FPLTQEFQRA LEMD Q LFINMQKN AINNK HELD RYTN LYNLNTIEK LQDNQKS
Hs  556 Q KIKS HSD-----E TSL GYF PNKKQLED L SKS---KE INQTRD LAKLN E ASSEQNKNHINN LERKEEQ
At  552 GVL GRLPPEKD-----MKRE QALRSI REYDDLSLKS---EAEKEVNMLQMKIQE NNS--LFKHNKDTESRKRYI

Sc  626 KEK IQLLSEN PEDCTIDYND LEETELSYKTALENLKM QTTLEEN AL IAERDSCCY CSRKEE--NESFKSKL
Hs  628 LSSYEDKLF VCGSQD ESL LR KEEIEKSS KQRAMLACATA SQFITQLTL ENQS--CCP CQRVFQTEAELQEA S
At  622 ESK QALKQES TIDAVPKLL SAK KRF DRK EYNMANGM Q EPFE ARQEHS----CPCC ERSET-ADEEASF K

Sc  704 LQE KT DF NFEK LKDT QNEKEYLHSLRLLEKHII NSFN-EK DNSQRCL KAK EETKTSRSKLDE EVDSTK K
Hs  706 DLQSKLFLAP LK TESE KKKF KRRDE LGLAP RQ IDLKEKEIFELRNKLQNVNRR QRLKND EQFT LG TIM
At  697 KQR KAS T LKALAVESSNA SVFQQLDKLRA FEEYSK LTTEIIFLAERKLQEHTEE GQKSEALD LG IS QIK

Sc  783 DEKELAESE RPLEKFTY EKELKD ENSSKTF SEELSIYNTS DG QTVDEL DQQRKMND LELRKTFSDLQF TK
Hs  786 PEEESAKVC TD-TIM RFQMELKD E KTAQQOAARLQG-IDLFT QQVFQEFQEKQ KLD SSKFELNRKLIQ QQ
At  777 ALK SIEALFQP-LENA RFQE VSYK QIEDIEYKLDFRGLGV T E SELSSLQSSKDEL GELEK RDDQ YMF

Sc  863 EK RENSRMINL KEKELTSE ESSTQKQN S RSKREN ID DSRVKLEARISLKNK EAQSVL K KNER
Hs  864 EQIQHLKSTTNEFKSEKLQSTNLQRRQQ---LEEQTVELSTE SLYR KAKEQ SPIETTLEKFQQK IN N
At  856 RDISCLQARWHA EEKAKA NILRD K---AEED ERLAEEKSQLDL VKYLTEA GP SKEKEQLLS YN SI RN

Sc  943 IQV NKQKT ADI RL RFQTIYNL VDFEA GFD LQTTIKLELNK------AQ LELKEQLLKSNE NEEKRK AD
Hs  941 TSN IAQ K NDIKEK VKN HGYMK TE HIQDG DLYMKQ ET K-----VIAQLSECEK RKENELRLM QD DT
At  933 QEYEELA KKR YQQEV AL LKASYKINDCFTRY DLKKGE LFQEKQRLSD QLQSCEA RNE AGE NRNKDL RN

Sc 1018 SN EE NLKQNLELIELKSQIQHI ESE SRL VQN -E ERDKYQEDS LRTRFE LSSENAGK LGE KQL NQIDSLT
Hs 1017 QKIQERWLQDNLTLRKRNEELKE EEGKQHLK -QMO LQ KSEHQKLEENI N KRNHNLALGRQK YEEEI F
At 1013 QD LRRN EDNLNYRTTKAKEE TREIESL EQ LNIG AAREAEIEK LRERELLSELNRCRGT SVYESSIS N

Sc 1097 LR-TDYKDIEKNYKEWE LQTRSFVTDD DYSKALDSAIMK HG KM IN IIDELW RTYSGTDIDTI IRSDE
Hs 1096 ELREPQ DAEE Y M TV TTE VNKDLD YYRTLDQAIMKFH MKMEEINKIIR LW STYRGQDI YIEIRSDA
At 1093 VEL QAQYKDI K HF Q IQL TTE ANKDLDRYYNALDKAL M FH MKMEEINKIIRELWQQTYRGQD DYI I SDS

Sc 1176 VS---S VKG SYNYRVVMYKQDVELDMRGRCSAGQKVLAS IIRLAL ETFGANCG IALDEPTTNLDEENIESLA L
Hs 1176 ENVSA DKRRNYNYRVVMLKGDTALDMRGRCSAGQKVLASLIIRLALAETFCLNCGIIALDEPTTNLDRENIESLAHAL
At 1173 G-----AGTRSYSY V MQTGDTEL MRGRCSAGQKVLASLIIRLALAETFCLNCGI ALDEPTTNLDGPNSESLAGAI

Sc 1253 HN IINMR HQ NFQLIVITHDE F GH NAAA T F VK DDRQKSQ EWVD NR T ---
Hs 1256 E IKSRSQQ NFQLFVITHDEDF E GRSE VEKFYR KKN IDQCSE VKCS SS GNVH
At 1248 RI EDR GQENFQLIVITHDE FAQ GQRQHA EK YRVAKDDM-----------------
```

Fig. 6

```
          1                                                    50
At    (1)  ----------------------MIGVDSKSSSTTFIETMVESEKTKHTC
Hs    (1)  --------------------------------MERKISR
Sc    (1)  MSQLTEFISCIPVVNEEQNEEDERGLCKIQIEDGAMLETLDENSLSGLRI
Consensus (1)                                   IETL E    SKIS

          51                                                  100
At   (28)  LRLEISGADPIFVKGTWHN---SRFDISVTDGSSS--------------
Hs    (8)  IHLVSEPSITHFLQVSWEKTLESGFVITLTDGHSA--------------
Sc   (51)  EKMLVSEGTGIFSKSSFGINDLRIFTGENIDEESKKYVWYELLKMLTGHK
Consensus (51)  IKLLIS A  IFLK SW      S F ISLTDG SA

          101                                                 150
At   (60)  --------------WICNATEEEVAERAAQWDQPVSEYL-KLAEQYLGFQ
Hs   (43)  --------------WTGTVSESEISQEADDMAMEKGKYVGELRKALLSGA
Sc  (101)  VYIASLDEKVVFTKWTCRMQDDEVWKVVMELES--SAIIRKIAELTLHPV
Consensus (101)        WTC MSEDEVA  A DLD   S YI KLAE  L

          151                                                 200
At   (95)  QPNSVYSFSDALEGSKRLSWTFEKEGTKLEWRWKCKPSDDSKKITVGILD
Hs   (79)  GPADVYTFNFSKESCYFFFEKNLKDVSFRLGSFNLEKVENPAEVIRELIC
Sc  (149)  KKGEIDLFEMADKLYKDICCVNDSYRNIKESDSSNRNRVEQLARERELLD
Consensus (151)  PADVYSF  A E  K I  NDKD S KE  F  K  DD  I RELLD

          201                                                 250
At  (145)  FLMEANIRLSEEVVNKTR-SFEKMRSEAERCLAQGEKLCDEKTEFESATY
Hs  (129)  YCLDTIAENQAKNEHLQK-ENERLLRDWNDVQGRFEKCVSAKEALETDLY
Sc  (199)  KLLETRDERTRAMMVTLLNEKKKKIRELHEILRQNNIKLSDDDVLDSALI
Consensus (201)  FLLET E S  MM  K E EKLIRE  DILAQ EK LSDKD LESALY

          251                                                 300
At  (194)  ----AKFLSVLNA--KKAKLR-------ALRDKEDSVRVVEE--------
Hs  (178)  ----KRFILVLNE--KKTKIR-------SLHNKLLNAAQEREKDIKQEGE
Sc  (249)  NTEVQKPISELNSPGKRMKRRKTVVEPQNLQKKLKDTSRRANRKISNQS
Consensus (251)      KFISVLNA KK KIR       AL KL  A  RE

          301                                                 350
At  (223)  -----EESTDKAESFESGRSDDEKSEEEASKKATSSKARGGKRAARS---
Hs  (215)  TAICSEMTADRDPVYDESTDEESENQTDLSGLASAAVSKDDSIISSLDVT
Sc  (299)  VIKMEDDDFDDFQFFGLSKRPIITAKDKLSEKYDDITSFGDDTQSISFES
Consensus (301)        EDS DK   FD SK DD A DDLS KASAA SKGD  S S  S

          351                                                 400
At  (265)  --------------------------------------------------
Hs  (265)  DIAPSRKRRQRMQRNLGTEPKMAPQENQLQEKEKPDSSLPETSKKEHISA
Sc  (349)  DSSSDVQKHLVSLEDNGIQISAGRSDEDYGDISGSESETDASAGEKKSSN
Consensus (351)  D A    K      G    A D    D   DS   SA      S

          401             423
At  (265)  ----------------------
Hs  (315)  ENMSLETLRNSSPEDLFDEI---
Sc  (399)  HSEQSGNDREPCLQTESETDIET
Consensus (401)           R       D
```

(54) Title: NUCLEIC ACID INTEGRATION IN EUKARYOTES

(57) Abstract: The invention relate to the field of molecular biology and cell biology. It particularly relates to methods to direct integration of a nucleic acid of interest towards homologous recombination and uses thereof. The present invention discloses factors involved in integration of a nucleic acid by illegitimate recombination which provides a method to direct integration of a nucleic acid of interest to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in a eukaryote with a preference for non-homologous recombination comprising steering an integration pathway towards homologous recombination. Furthermore, the invention provides a method to direct integration of a nucleic acid of interest to a sub-telomeric and/or telomeric region in a eukaryote with a preference for non-homologous recombination.

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7   C12N15/90   C12N15/82   C12N15/81

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7   C12N

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

EPO-Internal, WPI Data, PAJ, BIOSIS, MEDLINE, SEQUENCE SEARCH

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	BOULTON S J ET AL: "Components of the Ku-dependent non-homologous end-joining pathway are involved in telomeric length maintenance and telomeric silencing." EMBO (EUROPEAN MOLECULAR BIOLOGY ORGANIZATION) JOURNAL, vol. 17, no. 6, 16 March 1998 (1998-03-16), pages 1819-1828, XP002170161 ISSN: 0261-4189 the whole document --- -/--	1,2,4,7, 10,17, 19-21

[X] Further documents are listed in the continuation of box C.   [X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 June 2002	03/07/2002

| Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax: (+31-70) 340-3016 | Authorized officer

Oderwald, H |

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT		
Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	TSUKAMOTO Y ET AL: "Hdf1, a yeast Ku-protein homologue, is involved in illegitimate recombination, but not in homologous recombination." NUCLEIC ACIDS RESEARCH, vol. 24, no. 11, 1996, pages 2067-2072, XP002170162 ISSN: 0305-1048 the whole document ---	1,2,4,5, 10,17-21
X	WO 00 12716 A (PIONEER HI BRED INT) 9 March 2000 (2000-03-09) the whole document ---	1,3-5,8, 10,11, 17,19-21
X	BRESSAN D A ET AL: "The Mre11-Rad50-Xrs2 protein complex facilitates homologous recombination-based double-strand break repair in Saccharomyces cerevisiae." MOLECULAR AND CELLULAR BIOLOGY, vol. 19, no. 11, November 1999 (1999-11), pages 7681-7687, XP002170163 ISSN: 0270-7306 the whole document ---	1,2,4-7, 10,17, 19-21
X	TSUKAMOTO Y ET AL: "EFFECTS OF MUTATIONS OF RAD50, RAD51, RAD52, AND RELATED GENES ON ILLEGITIMATE RECOMBINATION IN SACCHAROMYCES CEREVISIAE" GENETICS,GENETICS SOCIETY OF AMERICA, AUSTIN, TX,US, vol. 142, February 1996 (1996-02), pages 383-391, XP000864573 ISSN: 0016-6731 the whole document ---	1,2,4-7, 10,17, 19-21
X	WO 00 68404 A (PIONEER HI BRED INT ;SHI JINRUI (US); MAHAJAN PRAMOD B (US)) 16 November 2000 (2000-11-16) the whole document ---	1,3-6, 8-11,17, 19-21
X	TEO S-H AND JACKSON S P: "Identification of Saccharomyces cerevisiae DNA ligase IV: Involvement in DNA double-strand break repair." EMBO (EUROPEAN MOLECULAR BIOLOGY ORGANIZATION) JOURNAL, vol. 16, no. 15, 1997, pages 4788-4795, XP002170164 ISSN: 0261-4189 the whole document ---	1,2,4, 10,17, 19-21

-/--

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	TSUKAMOTO Y ET AL: "Silencing factors participate in DNA repair and recombination in Saccharomyces cerevisiae." NATURE (LONDON), vol. 388, no. 6645, 1997, pages 900-903, XP002170165 ISSN: 0028-0836 the whole document	1,2,4, 10,17, 19-21
A	VERGUNST A C ET AL: "VirB/D4-dependent protein translocation from Agrobacterium into plant cells." SCIENCE (WASHINGTON D C), vol. 290, no. 5493, 2000, pages 979-982, XP002170166 ISSN: 0036-8075 cited in the application the whole document	12-21
A	BUNDOCK P ET AL: "TRANS-KINGDOM T-DNA TRANSFER FROM AGROBACTERIUM TUMEFACIENS TO SACCHOROMYCES CEREVISIAE" EMBO JOURNAL,GB,OXFORD UNIVERSITY PRESS, SURREY, vol. 14, no. 13, 3 July 1995 (1995-07-03), pages 3206-3214, XP000675831 ISSN: 0261-4189 cited in the application the whole document	
A	DATABASE EMBL 'Online! EBI; 3 March 2000 (2000-03-03) GALLEGO M E ET AL: "The Rad50 homolog of Arabidopsis" Database accession no. AF168748 XP002150323 abstract	
A	DATABASE EMBL 'Online! EBI; 1 November 1999 (1999-11-01) HARTUNG F AND PUCHTA H: "Isolation of the complete cDNA of the Mre11 homologue of Arabidopsis thaliana indicates conservation of DNA recombination mechanisms between plants and other eucaryotes" Database accession no. Q9XGM2 XP002170167 abstract	
A	WO 00 42205 A (NOVARTIS ERFIND VERWALT GMBH ;NOVARTIS AG (CH); BUDZISZEWSKI GREGO) 20 July 2000 (2000-07-20) the whole document	

—/—

International Application No

PCT/NL 01/00936

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
A	DATABASE EMBL 'Online! EBI; 1 October 2000 (2000-10-01) WEST C E ET AL.: "Arabidopsis thaliana ligase IV homologue is induced by gamma isrradiation and interacts with an Arabidopsis homolog of the double strand break repair protein XRCC4" Database accession no. Q9LL84 XP002170168 abstract	
P,X	VAN ATTIKUM HAICO ET AL: "Non-homologous end-joining proteins are required for Agrobacterium T-DNA integration." EMBO (EUROPEAN MOLECULAR BIOLOGY ORGANIZATION) JOURNAL, vol. 20, no. 22, 15 November 2001 (2001-11-15), pages 6550-6558, XP002201089 ISSN: 0261-4189 the whole document	1-11, 15-21

International Application No

PCT/NL 01/00936

Patent document cited in search report		Publication date	Patent family member(s)		Publication date
WO 0012716	A	09-03-2000	AU	5799599 A	21-03-2000
			EP	1108032 A2	20-06-2001
			WO	0012716 A2	09-03-2000
			US	6180850 B1	30-01-2001
WO 0068404	A	16-11-2000	AU	4975400 A	21-11-2000
			EP	1093523 A1	25-04-2001
			WO	0068404 A1	16-11-2000
WO 0042205	A	20-07-2000	AU	2290900 A	01-08-2000
			CN	1341151 T	20-03-2002
			WO	0042205 A2	20-07-2000
			EP	1141344 A2	10-10-2001

CORRECTED VERSION

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
4 July 2002 (04.07.2002)

PCT

(10) International Publication Number
WO 02/052026 A3

(51) International Patent Classification[7]: **C12N 15/90,** 15/82, 15/81

(21) International Application Number: PCT/NL01/00936

(22) International Filing Date:
21 December 2001 (21.12.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
00204693.6   22 December 2000 (22.12.2000)   EP

(71) Applicants *(for all designated States except US)*: **UNIVERSITEIT LEIDEN** [NL/NL]; Stationsweg 46, NL-2312 AV Leiden (NL). **STICHTING BINAIR VECTOR SYSTEEM** [NL/NL]; Floris Versterlaan 12, NL-2343 RS Oegstgeest (NL).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **HOOYKAAS, Paul, Jan, Jacob** [NL/NL]; Floris Versterlaan 12, NL-2343 RS Oegstgeest (NL). **VAN ATTIKUM, Haico** [NL/NL]; Boerhaavelaan 158, NL-2334 EV Leiden (NL). **BUNDOCK, Paul** [GB/NL]; Mijehof 345, NL-1106 HJ Amsterdam (NL).

(74) Agent: **PRINS, A., W.**; c/o Vereenigde, Nieuwe Parklaan 97, NL-2587 BN The Hague (NL).

(81) Designated States *(national)*: AE, AG, AL, AM, AT (utility model), AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA,
CH, CN, CO, CR, CU, CZ (utility model), CZ, DE (utility model), DE, DK (utility model), DK, DM, DZ, EC, EE (utility model), EE, ES, FI (utility model), FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK (utility model), SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— with international search report

(88) Date of publication of the international search report:
6 September 2002

(48) Date of publication of this corrected version:
27 November 2003

(15) Information about Correction:
see PCT Gazette No. 48/2003 of 27 November 2003, Section II

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

(54) Title: NUCLEIC ACID INTEGRATION IN EUKARYOTES

(57) Abstract: The invention relate to the field of molecular biology and cell biology. It particularly relates to methods to direct integration of a nucleic acid of interest towards homologous recombination and uses thereof. The present invention discloses factors involved in integration of a nucleic acid by illegitimate recombination which provides a method to direct integration of a nucleic acid of interest to a pre-determined site, whereby said nucleic acid has homology at or around the said pre-determined site, in a eukaryote with a preference for non-homologous recombination comprising steering an integration pathway towards homologous recombination. Furthermore, the invention provides a method to direct integration of a nucleic acid of interest to a sub-telomeric and/or telomeric region in a eukaryote with a preference for non-homologous recombination.